People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

University Abderrahemane Mira of Bejaia
Faculty of Exact Sciences
Computer Science Department

Thesis to obtain the Master of Computer Science
Degree in computer network administration and security
and artificial intelligence

Design and Production of Recommendation System for E-commerce

---

submitted by:

BENLAKEHAL Kinza **BENMAMAS Cylia**

Examination Committee:

chair: Pr. **TOUAZI Djoudi**
Examinateur: Pr. **ZIDANI Ferroudja**
supervisor: Pr. **EL BOUHISSI Houda**

September 2021

Acknowledgments

Praise be to Allah,

We would like to express our sincere thanks and gratitude to many people who offered us help and encouragements to the achievement of this paper.

At the top, it was an honour to be supervised by Professor EL-BOUHISSI Houda, who guided us throughout the realization of this thesis.special thanks goes to her for her serious professionalism, kindness, insightful suggestions and valuable pieces of advice.Your insightful feedback pushed us to sharpen our thinking and brought our work to a higher level. Thank you for everything.

And a special thanks to Benlekehal Fares for his guidance, respect and kindness.

We are also thankfull for the members of the examination committe, the honor they had done us by accepting to be the corrector of our dissertation. We thank you for agreeing to give your precious time to evaluate and criticize our work.

Finnaly, much thanks to our families and freinds, for their encouragement and support, which has allowed us to never give up.

Dedication

This study is wholeheartedly dedicated to all those who supported me. Firstly, I dedicate this work to my wonderful mother, who always encouraged me to do my best and follow my dreams, and to my father for his words of encouragement.

Additionally, this paper is dedicated to my lovely sisters; fazia, Elissia and Djamila for sharing every moment of our lives with each other.

To my amazing brothers; Faycel, Lekhel, El-Ghani,Massi and Zaki, the ones who I eternally thank for being the best brothers and for their support.

I also dedicate this work to my special one Rafik who believes in my dreams and for his endless help and always support me to never give up,thank you.

I would like to thank my lovely partner kinza, and my sweet friends Sabrina,tata ,Djamila and Sara .

.Cylia

Dedication

I dedicate this work:

To my lovely father, the one who sacrificed his life to raise me and be the person that I am today.

To my adorable mother and my idol, she taught me values, and how to be patient and strong to face my tough times.

To my soul mate sisters Selma and Massilya, my source of inspiration,they are always by my side.

To my perfect brothers, Fares and Amirouche that I found always in my side helping me and pushing me to do my best.

I would like to express much thanks and special gratitude to my husband Hani who always encourage me and pushes me toward success.

I extend dedication to all my family members; my unique grandmother, my grand father who always believe in my dreams and his endless supporting,aunts,and mu uniq uncle thank you for every thing ,my familly Saci , my perfect partner Cylia and also my friends;Sabrina, Sonia,Zahra,Tata, and Djamila who supported me from the beginning.

.Kinza

Abstract

Due to the tremendous expansion of world markets and financial transactions, the importance of electronic commerce has increased dramatically, in order to achieve a system of e-commerce be fully functional, reliable, efficient, secure and user-friendly, Proper system analysis and design procedures are essential. On the other hand, the tremendous growth of the Internet has completely changed the way most businesses work. The internet has created a way for e-commerce to provide businesses and customers with a unique way to sell and buy goods and services. When building an e-commerce website, there are several goals that need to be considered, one of them is how to increase the efficiency of the site to ensure customer participation and thus achieve the required material benefits. There are several methods which are followed to increase the efficiency of the site. The one of these methods is the recommender system. This thesis produces a design and realization of a recommendation system for the trade in using an example from Amazon.

Résumé

En raison de l'expansion considérable des marchés mondiaux et des transactions financières, l'importance du commerce électronique a considérablement augmenté, afin de parvenir à un système de commerce électronique pleinement fonctionnel, fiable, efficace, sécurisé convivial et l'analyse appropriée du système et les procédures de conception sont essentielles. D'un autre côté, l'énorme croissance d'Internet a complètement changé la façon dont la plupart des entreprises travaillent. L'Internet a créé un moyen pour le commerce électronique de fournir aux entreprises et aux clients un moyen unique de vendre et d'acheter des biens et des services. Lors de la création d'un site Web de commerce électronique, plusieurs objectifs doivent être pris en compte, l'un d'entre eux est de savoir comment augmenter l'efficacité du site pour assurer la participation des clients et ainsi obtenir les avantages matériels requis. Il existe plusieurs méthodes qui sont suivies pour augmenter l'efficacité du site ; L'une de ces méthodes est le système de recommandation. Ce mémoire produit une conception et une réalisation d'un système de recommandation pour le commerce en utilisant un exemple d'Amazon.

Contents

List of Figures

List of Tables

Nomenclature

RS Recommendation System

ML Machine Learning

CF Collaborative Filtering

CBF Content Based Filtering

KNN K-Nearest-Neighbors

Chapter I

General introduction

The continuous development in information and communication technologies have recently led to the emergence of distributed computing environments, which include several different sources of large volumes of data and several computing units. The most striking example of a distributed environment is the Internet, where more and more databases and data streams appear that deal with several fields, such as meteorology, oceanography, economics and others. Many researchers pay more attention to creating a suitable tool, which can help users get personalized resources. Recommendation systems (RS) are one such software tool used to help users get recommendations for invisible items based on their preferences.

Today, RS are present in the daily life of people and help them to make timely decisions, especially in electronic commerce(e-commerce), a reliable and efficient RS is essential. Sellers are quickly realized the power and importance to their own business.

RS have proven to be beneficial for both product sellers and users as they reduce the overhead required to find the items most likely to sell. As the recommendations are based on the tastes and interests of the users, it facilitates their decision-making process. The data required for RS comes from explicit user ratings after watching a movie or listening to a song, implicit search engine queries and purchase histories, or other knowledge about user /items and items/items. A large number of RS exist their objective is to filter and adapt the information for each user. The methods generally used for calculating the recommendation are either based on content or on the user's similarity to other users.

Many of the commerce Web sites are already using RS to help their customers find products to purchase. Amazon, is the one of the world's largest marketplaces, is said to have increased purchases by up to 35%, as McKinsey said. Another e-commerce giant, Alibaba, has also increased its purchases by 20% by implementing effective mechanisms to target users from RS.

For that in our thesis we present a generality on the RS and also a state of the art of the different approaches of the systems dedicated to the e-commerce, then we propose our design to realize our system by using an Amazon dataset that allows users to find their favorite products and according to their needs.

Problem formulation:

In recent years, new qualities of a good RS have been presented in the literature, in addition to the performance of predictions. An effective RS must offer new and diverse items to users, which meet their different interests and preferences, which requires the development of new ideas and techniques to formulate recommendations of interest. Thus, the interesting recommendations should contain various and relevant items taking into consideration the performance of the system. However, the e-commerce RS is not in-depth enough to analyze online consumer behavior in multi-source mining. RS only focus on product information and consumer purchasing behavior on the shopping platform, recommendation accuracy is limited for example if a customer searches Amazon for the least product. expensive and which corresponds to its categories it will be a bit slow and difficult. All these facts provided the motivation to propose a novel approach in building a RS to overcome the problems of information overload, data sparsity and cold start.

Objectives:

In our work, we are interested, in the use of the content-based filtering method and the use of two machine learning algorithms at the same time in order to be able to find the neighborhood of a user / item, to build a RS which can meet customers' needs and wishes.

Methodology:

Given that the problem of finding products that meet the needs of the customer, we present the approach adopted for our work by proposing a global methodological framework in order to meet the needs of the customers. Our work approach is based more precisely on the following steps:

Research and Analysis:

Which establishes a state of the art of the different technologies offered within the framework of RS with a comparative study.

Problem identification:

Which makes it possible to define the problem and the proposed solution.

System implementation:

Which sets up the proposed system, and how it works.

Thesis organization:

The present thesis is organized as follows:

- Chapter 1: In this chapter, we present the problematic and our objective and the plan of the work.
- Chapter 2: In this chapter we present a generality on RS, and its different types and approaches, and also the advantages and disadvantages.
- Chapter 3: This chapter is a state of the art of works that use recommendation systems for electronic commerce by giving a comparative study of these different approaches.
- Chapter 4: This chapter describes our approach, the architecture of the system, its design and also the algorithms proposed different steps followed for the construction of the system.
- Chapter 5: Used to finalize the presentation of our work. In particular, we explain the main functionalities of the tools that we have developed, as well as the evaluation. This chapter also groups together the different interfaces of the final application.
- Chapter 6: Concludes this this and opens up some research perspectives and future work.

Chapter II

General information on recommendation systems

Introduction

With the advent and ease of Internet access, we are increasingly exposed to a multitude of information. With all the blogs, newspapers, shops and more others, it brings a lot of diversity to users. Recommender systems, in the most general sense, are used to give the user of a site/function (whether that be a streaming video site like Netflix and YouTube, an online shop like Amazon or an online auction site like eBay) some added value. However, with this multitude of sources of information, overload information can become problematic. For example, a person who wishes to search for an article on the internet finds himself in front of a very large volume of proposals. This makes the task of choosing an article very difficult. Moreover, for facilitating decision-making for users, recommendation systems allow such filtering by offering them information according to their preferences in order to remedy this problem. In this chapter, we start by defining what is a recommender system and the fields of application, and its types of approaches, as well as their advantages and disadvantages. Then we present the steps to follow to build a recommendation system. Finally, we conclude the different problems of a recommendation system.

History of Recommendation Systems

The ability of computers to make recommendations to users was recognized early in the history of computing. Grundy [Rich, 1979], a library system, was a first step towards automatic recommendation systems(RSs). This system was quite primitive. He categorized users into "stereotypes" based on a short interview, and used those stereotypes to produce book recommendations. This work was an interesting first attempt in the field of recommender systems. However, its use has remained very limited. In the early 1990s, collaborative filtering appeared to be a solution to cope with information overload. The year 1992 saw the appearance of the Tapestry document recommendation system [Goldberg et al., 1992], as well as the creation of the GroupLens research laboratory, which worked explicitly on the problem of automatic recommendation within the framework of the forums of news from Usenet. Tapestry was intended to recommend to user groups material from newsgroups that might be of interest to them. The approach used was of the "nearest neighbors" type based on the user's history. We then speak of manual collaborative filtering, as a response to the need for tools for filtering information stated at the same time. The recommendation is the result of collaborative action by users who recommend documents to other users by assigning them interest scores according to certain criteria. Next, automatic collaborative filtering systems appear. GroupLens [Resnick et al., 1994] uses this technique to identify Usenet articles that may be of interest to a given user. Users should only assign grades or perform other observable operations (for example, read an article); the system then combines this data with the notes or actions of other users to provide personalized results. With these systems, users have no direct knowledge of the opinions of other users, nor of the articles in the system. In recent years, recommender systems have become a topic of increasing interest in the fields of human-machine interaction, machine learning, and information retrieval. In 1995, Ringo [Shardanand and Maes, 1995a], a music RS, based on user reviews and Bellcore, appeared successively. [Hill et al., 1995], a video RS. The same year, GroupLens creates the company Net Perceptions whose first customer was Amazon. Nowadays, RS have become essential components for most e-commerce sites [1].

II.1 Definition of a RS

The aim of the RS is to allow the user to quickly access content considered relevant by the system in relation to his research among the very large amount of information available on the Internet. Conversely, a RS does not receive a direct request from the user, but must offer him new possibilities by learning his preferences from his past behavior. A RS must therefore have access to a data history which can be in several forms: notes, purchases, clicks on web pages, browsing histories ... From this information, the RS will be able to tailor the response to the user. A RS must relate two entities: users and items. The items can be of various kinds (films, videos, restaurants, places of activity, etc.). The information that makes it possible to link these two entities is itself of different types: notes, purchases, clicks, history, etc. RSs mainly focus on the use of "notes" [Adomavicius and Tuzhilin, 2005]. The rating is generally located on a graduated scale, and it allows the user to express a positive or negative opinion on the item he is considering. In view of the scores he has already expressed on a set of items, the task of a RS has been mainly studied in two ways. The first is grade prediction, which aims to predict the grades a user would give to items they have not yet rated. The second is recommending an ordered list of items. Therefore, the items at the top of the list are therefore those that the system predicts as relevant for the user. The objective is both to minimize the time users spend on research, but also to suggest relevant resources that they would not have spontaneously consulted and thus increase their overall satisfaction [2][3].

II.2 Field of use

RSs are becoming essential in many areas, particularly in the areas of cultural industries. We can mention in particular:

- Cinema (Netflix, Movie lens),
- E-commerce (Amazon.com),
- Music (last FM),
- Tourism (Tripadvisor.com),
- Video on demand (Youtube.com).

II.3 Differents types of recommendation

II.3.1 Recommendation based on collaborative filtering

The principle of Collaborative Filtering (“CF”) is to use the evaluations made by users on certain documents (content), in order to recommend these same documents to other users, and without the need to analyze the content of documents. All users of the CF system can take advantage of the reviews of others by receiving recommendations for which the closest users have made a favorable value judgment and this without the system having a content extraction process documents. Thanks to its independence with respect to the representation of data, this technique can be applied in contexts where the content is either unavailable or difficult to analyze, and in particular, it can be used for any type of data: text, image, audio and video. In addition, the user is able to discover various interesting areas, because the principle of collaborative filtering is not based on the thematic dimension of profiles, and is not subject to the "funnel" effect. Another advantage of CF is that users' value judgments incorporate not only the thematic dimension but also other factors relating to the quality of documents such as diversity, novelty, etc. The CF suffers from several big problems. The main problem being the cold start: it is the fact that a user has to vote on a lot of object before getting the recommendations [5].

Collaborative filtering process

The CF process follows the steps given below:

- **Assessment of recommendations**

According to the basic principle of CF, users must provide their ratings on documents so that the system trains communities. Evaluating a recommendation can be done explicitly or implicitly, as follows:

.Explicit: The user gives a numerical value on a given scale (for example from 1 to 5, or from 1 to 10, etc.), or alternatively, a qualitative value of satisfaction, for example, bad, average, good and excellent.

.Implicit::The system induces user satisfaction through his actions. For example, the system will judge that a deleted recommendation corresponds to a bad rating, while a printed or saved recommendation can be interpreted as a good evaluation.

Community training:

The process of forming communities is the core of a CF system. For each user, the system must calculate their community; usually this is done by the proximity of user ratings. To do this, we can first calculate the proximity between a given user and all the others. Then, and in order to create the community of the user, unlike the nearest neighbors method, using a threshold for the proximity level or a threshold for the maximum size of the community, due to its performance and its precision.

Production of recommendations

In these latter processes, once the user's community is created, the system predicts the interest that a particular document may have for the user based on the reviews that community members have made on the same document. When the predicted interest exceeds a certain threshold, the system recommends the document to the user.

Profiles and communities

Here, we discuss profiles based on user rating history, as well as communities, which are the two key factors of SFC. The problem of information overload can be addressed by customizing access to information, using profiles representing relatively stable interests of users. In other words, user profiles are used as persistent criteria in the search for information.

- **User profile:**

The user profile is made up of weighted predicates. The weight of a predicate expresses its relative interest for the user. It is specified by a real number between zero and one. The profile is gradually enriched as the user evaluates the documents received. In addition to basic identifying information (for example, identifier or civil status elements), the user profile can group a wide variety of information as needed. Among these, we can cite:

- Personal characteristics that can strongly influence the interaction (age, sex, etc.).

- The interests and general preferences of the user relating to the task at hand, which allow adaptation to his expectations.

- Quality. This dimension contains all the factors reflecting preferences regarding the quality of information, such as data availability, conciseness, style and structure of the document, etc. In this dimension, we are particularly interested in information diversity.

- Security. The security dimension in the context of collaborative filtering is the level of confidentiality regarding all other criteria.

- A history of interactions with the service, which can be used to model behavioral habits.

- **Communities**

The notion of community in a collaborative filtering system is defined as the grouping of users according to the history of their evaluations, so that the system calculates recommendations. From this perspective, profiles are an interactive factor, while communities are seen as an internal factor of the system.

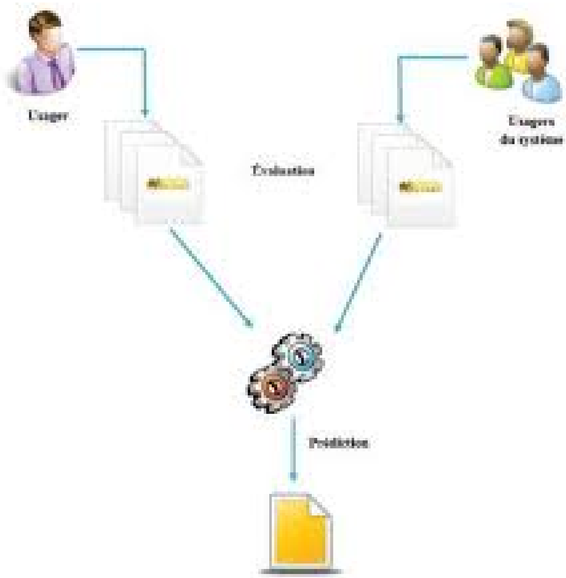


Figure II.1: Recommendation based on collaborative filtering.[1]

II.3.2 Content-based recommendation

CBF, which is a general evolution of studies on information filtering, is based on the content of documents (topics covered) to compare them to a profile itself made up of themes. Each user of the system then has a profile, which describes areas of interest. For example, the profile may contain a list of themes or preferences that the user likes or dislikes. When a new document arrives, the system compares the document description with the user's profile to predict the usefulness of this document for that user.
The advantage of cognitive filtering systems, based on content, is that they allow documents to be associated with a user profile. In particular, by using indexing and artificial intelligence techniques. The user is independent from others that allows him to have recommendations even if he is the only user of the system. In order for example to recommend films to a user, the system analyzes the correlations between these films and the films previously viewed by this user. These correlations are evaluated by considering attributes like title and gender. Therefore, among these films, those, which will be recommended to the user, are the most similar (in terms of attribute) to the films viewed by this user [5].

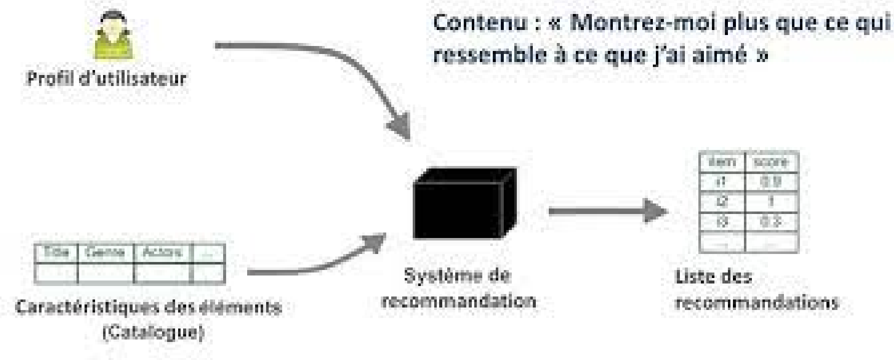


Figure II.2: Content based filtering recommendation[1]

II.3.3 Item descriptor and user profile

The content-based filtering algorithm can perform the matching between a content descriptor (such as documents, books, etc.) and a user profile and determines the degree of relevance of each article (or content) for the users. Potentials. If many articles accumulate in a certain amount of time, the content filtering algorithm can order the articles according to their relevance to each of the potential users.

Content representation - the article descriptor

An article descriptor consists of a set of concepts that can be represented by a domain ontology. The concepts that represent an element are the most specialized in a branch of the hierarchy. Obviously, an article can be represented with many ontology concepts, each concept can appear in any branch of the ontology hierarchy and at any level, and it depends on the actual content of that article. It should be noted that the profile could include sibling concepts, that is to say children of the same concept.

Representation of users - the user profile:

A user profile content-based consists of a weighted list of ontology concepts, representing its preferences (interests). Obviously, the user's profile can have many ontology concepts, each one appearing in different branches and different levels of the hierarchy. For example, the user's profile might include only "sport", or "sport" and "soccer", or "soccer" and "basketball", or all three - in addition to many other concepts. This means that a certain concept in an article descriptor can be compared with more than one equivalent concept in the user's profile.

Similarities between an article descriptor and a user profile

An article descriptor and a user profile are similar to some degree if their profiles include common concepts (the same) or related concepts, that is, concepts that have some kind of father-son relationship. An article descriptor and a user profile can have many common or related concepts; obviously, the more common or relative the concepts, the stronger their similarity. For example, if the user's profile includes "football" and "sport", that profile is similar (to some degree) to an article that includes these two concepts, but is less like an article that includes just "sport.", and it's more like an article, including “sport”and“football”.

Hybrid filtering

Many systems rely on the combination of the CF approach and the CB approach, making them so-called filtering systems Hybrids. In general, hybridization is carried out in two phases: (i) separately apply Collaborative filtering and other filtering techniques to generate candidate recommendations, and (ii) combine these sets of preliminary recommendations according to certain methods such as weighting, mixing, cascading, switching, etc., in order to produce the final recommendations for users [Nguyen, 2006]. [6]
More generally, hybrid systems manage content-oriented user profiles, and the comparison between these profiles gives rise to the formation of user communities allowing collaborative filtering. [Burke, 2002] has made the best description of hybrid methods. Therefore, according to Burke we can distinguish seven ways of combining traditional methods:

- **Weighted**

A hybrid method that combines the output of separate approaches, using, for example, a linear combination of scores from each recommendation technique.

- **Switching**

A technique allows you to choose a recommendation model among several based on several criteria. Determining the appropriate technique depends on the situation. The system must then define the switching criteria or the cases where the use of another technique is recommended. This allows the system to know the strengths and weaknesses of the recommendation techniques that constitute it.

- **Mixed media**

In this approach, the recommender does not combine, but increases the description of the datasets, taking into account user estimates and item descriptions. The new prediction function has to deal with both types of descriptions and avoids the problems posed by CF, namely, cold start. Features combination in a hybrid based on the combination of features, data from collaborative techniques is treated as a feature, and a content-based approach is used on this data.

- **Waterfall**

The waterfall involves a systematic process. In this case, a recommendation technique is applied first, producing a set of potential candidates. Then, a second technique refines the results obtained in the first step.
The advantage of this method is that if the first technique generates few recommendations, or if these recommendations are ordered to allow rapid selection, the second technique will no longer be used. Feature augmentation the feature augmentation is similar to the cascade, but in this case the results obtained (ranking or classification) of the first technique are used by the second as an added feature.

- **Meta-level**

In a meta-level-based hybrid, a first technique is used, but differently than the previous method (increase in characteristics), not to produce new characteristics, but to produce a model. In addition, in the second step, the entire model will be used as input for the second technique [Arnautu, 2012].

II.3.4 Demographic recommendation

It is a simple recommendation that offers items in relation to the demographic profile of the user. It involves dividing users into several classes or groups in relation to skin information such as sex, age, profession, location, language, country, etc. The principle of this approach is that two users who have evolved in a similar environment share common tastes as two users who have evolved in different environments and therefore do not share the same codes [Bouchindhomme and Rochlitz, 1992]. Many sites use this simple solution to provide a "personalized" content offering. For example, users are redirected to a particular website based on their language or country. These approaches have been very popular in the marketing literature, but have received little attention in the field of recommendation algorithms[7].

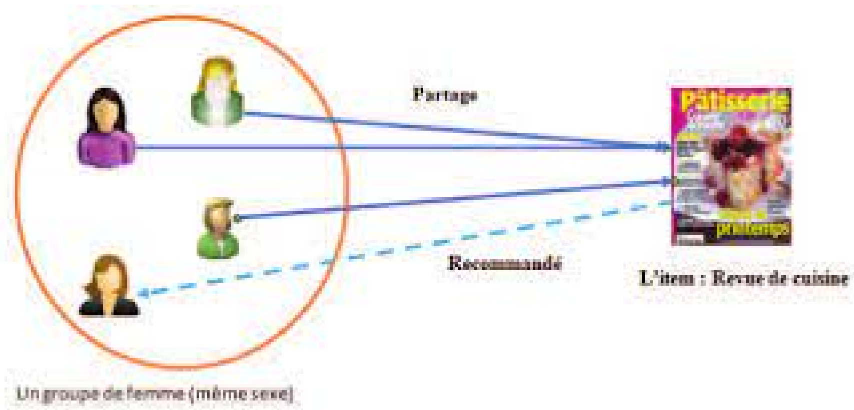


Figure II.3: Demographic recommendation [1].

II.3.5 Knowledge-based recommendations

Recommendations are generated using specific knowledge, certain item characteristics of which meet the user's preferences. Generally Systems have more reliable knowledge based compared to other types of recommendation if limited data is available, ie, if the system cannot rely on the existence of a user history. However, if the knowledge system is not designed to learn notes or user actions[7].

Case-based reasoning

Case-based reasoning takes advantage of real-world regularity to solve problems by seeking the solution of a similar case encountered and solved in the past. [Piamrat et al., 2009] have used this approach in RS, they estimate how much the user's needs or preferences (problem description) correspond to possible recommendations (problem solutions) based on behavior previous consumption (previous cases).

Constraint-based reasoning

Another type of knowledge-based system is a constraint-based recommendation. Constraint-based recommendation leverages predefined knowledge bases that contain explicit rules on how to relate customer requirements with item functionality. For example, a user may be interested in purchasing products with a certain set of characteristics and in a specific price range.

II.3.6 Community recommendation

Community or social recommendation, since most social networks (Facebook, twitter, etc.) rely on this classification in their recommendations (figure 2.4). The basic idea is therefore to say that if users have shared the same interests in the past, there is a good chance that they will also share the same tastes in the future. The system offers recommendations based on the user's relationships with these friends in the social network, and sometimes this recommendation also depends on the user's trust value in each of their friends, the best-known example of this recommendation is the pages and groups section that appears on the right side of a Facebook page. The decisional importance of the button «I Like» of Facebook has given increasing success with 55% of users being influenced by their friends. [7].

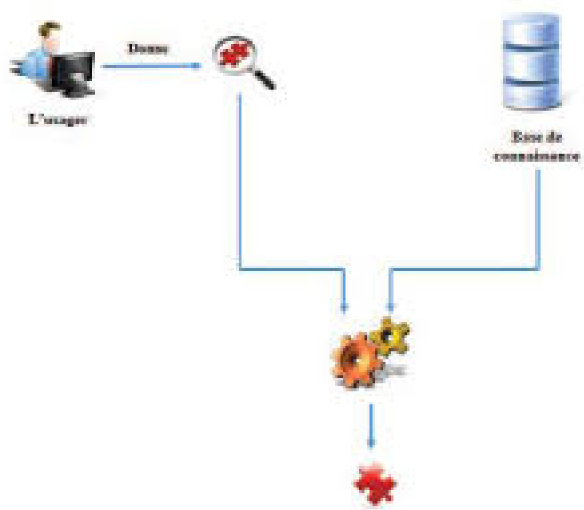


Figure II.4: Community recommendation [1].

II.4 Advantages and disadvantages of RSs

Like all computer systems, the RS has advantages and disadvantages that we will summarize them in the following table:

Techniques	**Benefits**	**Desadventages**
Content-based filtering	•No need for a large community of users to be able to make. •A list of recommendations can be generated even if there is only one user. •Quality grows over time. •No need for information about other users. •Take into consideration the unique tastes of users.	•Content analysis is required to make a recommendation. • Problem of recommending images and videos in the absence of metadata. •Requires user profile.
Collaborative filtering	•Does not require any knowledge about the content of the item or its semantics. •The quality of the recommendation can be assessed. •The greater the number of users, the better the recommendation.	•Cold start. •New Item. •New user. •Confidentiality issue. •Complexity: in systems with a large number of items and users, the computation grows linearly.

Table II.1: Advantages and disadvantages of RS[1].

II.5 Problems with RSs

II.5.1 Critical mass

This aspect illustrates the difficulty of managing the fact that; there are a few articles actually evaluated, or a few users who carry out these evaluations. As a result, the RSs has little or no data to generate these predictions. Thus, it is necessary to exceed a sufficient number of evaluations before which the recommendations, provided by the system, are not relevant.

II.5.2 Cold start

Often, we find ourselves faced with the problem that one user cannot be compared to any other. This is because few or no users have rated a given article, or a given user has rated very few or no articles.

II.5.3 Induction principle

RSs are based on the principle that a user who exhibited behavior in the past will tend to exhibit similar behavior in the future. However, this principle is not necessarily valid in the real context. Indeed, a user can completely change his domain of interest or have several. To deal with this problem, techniques of interest drift or context shifts have emerged.

II.5.4 Security and credibility

Referral systems cannot prevent acts of deception. This makes it easy to forge a new identity and engage in vandalism, such as providing bad information to the system. It should be noted that in distributed systems without central authorities, it is more difficult to control the identity of users and to penalize malicious behavior.

II.5.5 Collection of preferences

One of the most important and difficult steps in recommender systems is collecting user preferences. Indeed, obtaining feedback from users on a given resource that they liked, liked less, or not at all more, is a difficult task. Thus, techniques for collecting user preferences, intrusive or not, have emerged.

II.5.6 Computational complexity

Evaluating similarities, an integral part of collaboration filtering, involves a few processes with a lot of computation. For a large number of users, the calculation of the similarity for all the individuals in the neighborhood becomes infeasible. Thus, good performance can only be ensured by limiting these calculations. To do this, it is important to sufficiently restrict the size of the community. Thus, intelligent filtering mechanisms are necessary that do not sacrifice too much important information.

II.5.7 Protection of private life

Another problem that affects RSs is the protection of sensitive information constituting the user profile (personal information, interests, tastes, habits, etc.). Given the nature of the information, these systems must provide such protection. Thus, means are necessary to preserve the anonymity of users and encrypt the data transmitted[1].

Conclusion

We have attempted throughout this chapter to focus on recommender systems, as well as their main types and their advantages and disadvantages. A point on the problems of RSs was discussed and the areas of application of this system, and among these areas including e-commerce.
In the next chapter, we will present the state of the art on RSs for e-commerce.

Chapter III

State of the art

Introduction

With the rapid development of the Internet and the continuous expansion of the scale of e-commerce, the number and variety of products is increasing rapidly. Merchants offer many products through shopping websites, and usually customers spend a long time to find their products. Browsing a lot of information and irrelevant products will scare consumers away due to information overload. In the age of e-commerce, users need an electronic shopping assistant, which can recommend interesting or satisfying products based on their interests and hobbies. To solve all these problems, a recommender system is needed. Many approaches have been proposed to build accurate and efficient recommender systems. Automatic recommender systems use a wide range of techniques. In this chapter, we will present a state of the art of the main related works according to e-commerce recommendation. A comparison study of the proposed approaches, helped us to identify the main issues and propose solution to remedy.

III.1 RS for E-commerce

The rapid growth of emerging technologies and the internet has shifted the world to an electronic world where most things are digitized and available with just a mouse's click. Most of the business transactions are done over the internet using online shopping which makes e-commerce more popular. E-commerce is very popular these days. Customers bought more and more products on the web, and businesses are selling more and more products on the web. Whenever a user wants to buy a product on the web, he visits an online store and searches the item that interests them. There are many popular e-commerce sites like ebay.com and amazon.com. These online stores sell many items. For a single item, many brands and models are

available. The ability for the customer to choose from a large number of products increases the information-processing load before he decides which products meet his needs. If the customer is not certain to choose product, he may face the problem of information overload. He may come across a situation where he may be unable to decide which product to purchase. RSs have a particularly important place in online marketing. Thanks to them, e-commerce companies have been able to differentiate themselves from their competitors, make life easier for current customers and reach their potential customers. Every time a user visits a site and selects a product to buy, the sites recommend other products to purchase. Product RSs attempt to predict which products may interest the user, based on product information and user profile. Depending on the companies' strategies, numerous recommendation techniques are combined to meet the user needs. These methods have different advantages and disadvantages, and cannot resolve all the problems stated before. Usually, companies use several approaches and combine them to have a better recommendation, evaluated by certain predefined criteria in their context as well as their objectives.
Most existing RSs use CF or content-based or hybrid methods that combine these two techniques. If the recommendation process is quite simple, its implementation is however a complex task. These difficulties lie in certain aspects such as the collection and selection of relevant data, the size and quality of data, the scarcity of data, the construction of user profiles, the prediction for new user profiles or new products [8][9].

III.2 Literature review

E-commerce RSs are relatively a recent domain whose state of the art has been studied different works. The research challenges are related to collecting user information. This section describes the main related works according to the e-commerce field and summarizes their main features, advantages and disadvantages. Proposals are categorized across a range of research areas, including code's regarding approaches based on collaborative filtering, on ontologies, based on ML and hybrids, which use ontologies as a semantic model and ML algorithms for prediction.

III.2.1 Hybrid RSs

Hansson [10] proposed a new hybrid approach to product recommendation which involves the content-based clustering and connections between clusters using collaborative filtering to make good product recommendations. The goal is to develop a hybrid model capable of making good product recommendations based on product information and behavioural data by clustering on that data, i.e., putting products together in groups based on how similar they are, and then recommending items The authors use clustering algorithm on two specific attributes, categorical and numerical attributes including k-Means, DBSCAN and hierarchical clustering. The algorithm was tested on real product and purchase data from two different companies - a big online book store and a smaller online clothing store. It is evaluated both for functionality as a back filler to other algorithms and as a strong individual algorithm. The evaluation mainly looks at the number of purchases as metric but also uses accuracy and recall as evaluation met-rics. The algorithm shows some promise for using it as an individual algorithm. As the time for this proposal is limited, this work focuses only on non-personalized recommendations.

The authors in [11] present a proposition to utilize flexible neural network architecture called Deep Hybrid Collaborative Filtering with Content (DHCF) as a product recommendation engine. The system was tested on 2018 Amazon Reviews Dataset, using repeated cross validation and compared with other approaches: collaborative filtering (CF) and deep collaborative filtering (DCF) in terms of mean squared error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). DCF and DHCF were proved significantly better than the CF. The experiment shows that DHCF can outperform other approaches considered in the study, with more robust scores.

III.2.2 RSs based on ML

Incorporating ML algorithms into RS will certainly improve the accuracy of recommendations. In this section, we present the main RSs that use ML algorithms.

The authors in [12] proposed a model-based collaborative filtering recommender systems based on probabilistic. To overcome the repetitive recommendation, this paper proposed a model-based collaborative filtering approach. This approach uses three features, unique identity of the user, query that user searched and the time difference between the time at which user fired the query and the time at which user clicked the item from the search results. The approach accounts for various features of user behavior on e-commerce sites and builds a training model by extending the Naive Bayes approach, and improving it using the bigram language model. The imports datasets for the experiments include user behavior on an e-commerce site ranging over 3 months, 1.8 million bought products and 1.2 million customers. The proposal uses Naive Bayes algorithms with bigram language model to improve search query analysis. Therefore, search query, click time and query time are used as features for Naive Bayes algorithm model. The proposed system predicts the probability of products and products will be recommended to the user to make top-N recommendations. In extending the Naive Bayes approach, and improving it using the bigram language model. They concluded that this system improved probabilistic model has 14% higher value of F1 measure as compared to simple Naive Bayes model, and improves search query analysis.

Liu et al. [13] analyze and study the shortcomings and challenges of traditional online shopping behavior prediction methods, and propose a network shopping behavior analysis and prediction system. Through the analysis of customer behavior data, the system obtains the customer, purchase behavior rules included in the customer, and stores the discovered rule knowledge in the knowledge base. Selects linear model logistic regression and decision tree based XGBoost model. After optimizing the model, they found that the nonlinear model could make better use of these features and get better prediction results. They used many methods which are, E-commerce user behavior prediction model based on decision tree algorithm, Prediction model of commodity purchase behavior based on XGBoost method, E-commerce sales forecasting model based on machine learning algorithm, and stable volatility model. This new system simplifies the complexity of the model to a certain extent and improves the classification accuracy of machine learning. The nonlinear model can make better use of these features and get better prediction results. They have used java language for the data mining process, and MySQL for data preprocessing and feature extraction. The logistic regression model uses language modeling and evaluation models. SVM, decision trees, and XGBoost use Python to build models and optimize models.

III.2.3 RSs based on collaborative filtering

Different recommendation approaches in the e-commerce domain based on collaborative filtering, presenting us with some RSs that use this approach.

Zhang et al. [14] propose an improved CF- recommendation algorithm, NewRec, and verify the feasibility of NewRec algorithm through experiment simulation, proving that it can improve the recommendation quality of E-commerce recommendation system. In order to better solve the problems of data sparsity and rating time factor, they adopt level filling method to predict the nonrated items and finally combines time weights in the recommendation prediction stage to improve the recommendation accuracy of the algorithm, they use Hierarchical Filtering Method and Improvement of Recommendation Timeliness. They found than their algorithm is better than the traditional collaborative filtering recommendation algorithm in running efficiency and recommendation accuracy. This algorithm can improve the recommendation quality of E-commerce RS, to predict the nonrated items and finally combines time weights in the recommendation, data sparsity in user-item rating matrix and timeliness of user evaluation. In the other side, they concluded than this algorithm has poor expansibility of preference models, inability to adapt to dynamic change of datasets, resulting in lack of time information that can be used, and inability to solve cold start problems very well. At present, there are still many problems and shortcomings in the studies of Ecommerce personalized recommendation. For user personalized recommendation, the improved collaborative filtering algorithm in this paper fails to consider the influences of context and user interaction behaviors, which need further thorough studies in the future.

Another proposal [15] which proposed an e-commerce recommender system based on collaborative filtering using a data mining approach. In this study, basic and common limitations of RS in e-commerce concerning cold start, scalability, sparse user-item matrix, and the change in user interests have been considered with the objective to resolve or mitigate these limitations. Accordingly, using a set of data mining methods, a comprehensive e-commerce recommender system has been proposed for overcoming these limitations. Considering the results of the study, the performance of the proposed system is better than traditional CF systems. The general results of the study indicate that the accuracy of the new recommender system has increased significantly compared to the traditional system. Therefore, the accuracy of the proposed system in the recommendation of 30 items has been about 28% higher than the traditional one, in this study, the two-stage clustering method is used for segmentation by combining SOM (determining the optimal number of clusters) and K-means (creating clusters) methods. They concluded that this system used for improving the quality of the decision making process in EC environments, since they help online decision makers by providing necessary information and useful recommendations to select items consistent with their desires and requirements.

Jiang et al. [16] present their proposal which aims to remedy to the problem of low accuracy of the traditional slope one algorithm and the untrusted ratings in recommender systems. Moreover, they propose a slope one algorithm based on the fusion of trusted data and user similarity. The algorithm we proposed can applied in many applications, such as the RS for social networks. They implement their experiment on parts of Amazon's items rating dataset; they do the evaluation in four aspects. Firstly, they compared slope one algorithm based on trusted data and the traditional algorithm. Secondly, they researched the difference between the slope one algorithm based on the fusion of trusted data and similarity and the algorithm based on trusted data using MAE. Thirdly, they comparison between the slope one algorithm based on the fusion of trusted data and similarity and the algorithm based on trusted data using RMSE. Finally, they had a comparison between slope one algorithms based on user similarity under different sizes of datasets. Therefore, they concluded that their algorithm performs more accurately than the traditional slope one algorithm.

III.2.4 RS based on a hybrid ontology

Guia et al [17] presens a new hybrid approach that combines the simplicity of collaborative filtering with the efficiency of the ontology-based recommenders. The experimental evaluation demonstrates that the proposed approach presents higher quality recommendations when compared to collaborative filtering. The main improvement is verified on the results regarding the products. They propose an innovative RS approach, which combines the simplicity of the most prevalent algorithm in collaborative filtering, which is the K-Nearest Neighbor algorithm (KNN) with the efficiency of ontology-based recommenders. The experimental evaluation shows that our hybrid ontology-based system can recommend, in comparison with the collaborative filtering version, products that match user preferences more easily and belong to new categories for the active user. They concluded that with ontology-based models finding users that have similar preferences with the active user is simpler. Nevertheless, the time to apply the KNN to find the k-nearest Products is too high.

The authors in [18] describe an approach to personalize the e-service proposal in a multi-service platform by combining the results discovery methods, based on the semantic description of the e-service, with information from the user profile. The aim of the proposal is to present a model capable of recommending products and services likely to be of interest to users by combining semantic discovery methods with user modeling techniques. They proposed the OWL-S ontology for e-services description and the Good Relations ontology for products/services offering, and for the description of their commercial features. Their proposal is to use semantic technologies and Linked Data principles to address the required flexibility of the aforementioned model; database proposed in linkeddata.org/.then, this approach is used to improve the service discovery, thus making it easier to offer solutions with respect to user interests.

III.3 Analysis and comparison

The table below represents the main characteristics of the different column approaches:

• The column «approach» defines the author of the proposed approach.

• The column «Categories of the approach» Denotes the category of recommender system used under the approach.

• The column «Data source » Indicates the set of data used to manage the system.

• The column «used technique » Defines the different methods used in this approach.

• The column « supported tools »Defines the software used in this approach.

• The column « advantages » Presents the main advantages of this approach.

• The column « disadvantages » Presents the drawbacks of this approach.

Systems set above use different datasets, different results in the same e-commerce domain. It represents a huge opportunity for trade electronic, working with large amounts of data and introducing sophisticated algorithms, online stores online should see their conversions increase with current referral systems.

Most of these approaches using the Collaborative filtering techniques; this method attempts to find a group of users who have the same tastes. In addition, preferences of the target user. For hybrid recommendation, it incorporates two or more recommendation techniques to limit the weaknesses of each. However, the use of recommender systems has exposed many challenges: has data scarcity, cold start issues, fraud and privacy in some areas like healthcare.

Some systems are based on the collaborative content-based techniques; this type of RS is based ones on profiles. In fact, we build profiles for users and as well as for products. Other articles have particular methods of eliciting user interests and preferences based on ontologies as a model. Semantics. The latter may solve the cold start issue due to an initial lack of reviews for new users.

Most of the systems cited involve machine learning algorithms coven the KNN algorithm, K-means (creating clusters methods), the Naive Bayes algorithm, these algorithms are based on estimating the likelihood that a person likes a certain product. To calculate this probability,

Category	**Approach**	**Data source**	**Used technique**	**Supported tools**	**Advantages**	**Disadvantages**
Hybrid based recommender systems.	[10]	Apptus' customers. product catalog and event files.	Machine Learning algorithms. Clustering Collaborative content and filtering	Yes.	k-Means algorithm are that it is fast, simple and can process large data sets efficiently.	Finds only a local minimum and it can generate bad clustering Needs the number of clusters it should divide the points into; it cannot compute the best possible value for k by itself.

Hybrid based recommender systems.	[11].	Amazon Reviews Dataset 2018.	Collaborative Filtering. Deep Collaborative Filtering. Deep Hybrid Collaborative Filtering.	Yes	More flexible and extensible, which in consequence leads to statistically better generalization results	
ML based recommender system	[12]	User behavior on an e-commerce site ranging	Naïve Bayes algorithms. Bigram approach. Collaborative filtering.	YES	Find the highly probable product based on user's query.Improve search query analysis	
ML based recommender systems.	[13]	Customer purchase behavior	decision tree algorithm Prediction. XGBoost method.	Yes.	Simplifies the complexity of the model Improves the classification accuracy of machine learning.	

					Nonlinear model makes better use of these features and get better prediction results.	
Collaborative filtering	[14]	movielens.org	Hierarchical Filling Method, Improvement of Recommendation Timeliness.	Yes	Improve the recommendation quality of E-commerce RS Predict the nonrated items and finally combines time weights in the recommendation	Poor expansibility of preference models. Inability to adapt to dynamic change of dataset Resulting in lack of time information that can be used, and inability to solve cold start problems very well

Collaborative filtering	[15]	Website in Iran, which involves 2,542 customers and 17,000 records.	K-means.	Yes.	Improving the quality of the decision making process in EC environments.	
Collaborative filtering	[16]	Amazon's items rating.	User similarity.	Yes	Works efficiently and effectively	
Ontologies based recommender system.	[17]	Amazon reviews between May 1996 and July 2014.	K-Nearest Neighbor algorithm Owl-s ontology.	Yes	Simplicity to find users that have similar preferences with the active user.	The time to apply the KNN to find the k-nearest Products is too high.
Ontology-based Recommender system.	[18]	Linked Data Ontologies.	OWL-S ontology	Yes	To improve the service discovery.	

Table III.1: State of the art of the main contributions relating to e-commerce recommendation.

there are different ways to understand the problem. The clustering approach is widely used to generate the cluster because it is a powerful unsupervised learning method to properly assess the large amount of data created by applications. However, the recommendation will only be better if the clusters formed are good.

According to this analysis, all the selected mechanisms are compared based on certain crucial metrics such as security, response time, progressivity, accuracy, and cost of operation, implicit / explicit data source and independence. The results confirmed that most studies aim to improve the accuracy of recommendations, but that the safety, response time, novelty, diversity and serenity are not taken into account in many articles. To this end, we would be able to choose the best ML algorithm that can handle the characteristics of RS. In addition, the combination of an RS based on ontology and ML algorithms is a promising approach to improve the accuracy of recommendations. However, to our knowledge, little corresponding research has been carried out based on ontologies and ML algorithms.

From this study, and the problems we have mentioned, we offer a new approach based on ML algorithms, which used to identify user needs and provide precise and effective recommendations. Inspired content based filtering approaches, our method which combines linear Kernel algorithm-based recommendations with a KNN algorithms provide more diverse recommendations after processing the data and categorizing it. One of our main goals is to provide recommendations for similar products and rank them in order of growth by price.

III.4 Conclusion

In this chapter, we have gone through our literature review and related work in the field of RS. We introduced two important domains for building effective recommendation models : ontology and machine learning. We defined what machine learning is, enumerated its types and enumerated some of its algorithms used to build RS. The same work was done for ontologies where we defined the concept and the benefit of using it in RS.

In the next chapter, we will introduce our approach in building our recommendation systems by describing the different steps followed.

Chapter IV

Recommendation System for E-commerce

Introduction

The replacement of traditional shopping fashion by the various modes of online shopping in real-time. Because of traditional shopping, most of them are getting into real feel about the product whichever they buy. The product features will be manually realized by the consumers whereas in online shopping all the consumers believe the descriptive summary of the products and the various factors based on the sold historical data.

Now a day's modern shopping method is moving gradually towards hitting a greater number of customers. Here recommendation system playing a vital role in suggesting the product by considering the earlier records and increasing the demand. Many of the consumers are attracted by factors like deals on an item, rating, review, and cost of the product. Through these factors, most of the consumers are attracted to taking online shopping instead of traditional shopping methods. For suggesting the products to consumers, many kinds of recommendation algorithms are applied using machine learning and deep learning technology to train the system automatically by observing the customer behavior patterns. But the believing factors of the product will be forged some time; in such cases, consumers are not satisfied with their expectations.

Over the years, SR have been widely used in e-commerce sites but they still pose research and practical challenges, including scalability, richer data, consumer-centric recommendations, anonymous users, and login recommendations. They are used in large sites such as Amazon, where millions of products are vended, actively making recommendations to thousands of users simultaneously in real time. The observed performance includes latency in generating

recommendations, number of requests being processed, number of consumers, number of products, and a large amount of rating and review data. Recommendations are also updated due to the nature of the search history change.
In this chapter, we present our RS according to e-commerce and its main steps.

IV.1 Problem formulation

RSs are changing from novelties used by a few E-commerce sites, to serious business tools that are re-shaping the world of E-commerce. Many of the largest commerce Web sites are already using RSs to help their customers find products to purchase. A RS learns from a customer and recommends products that she will find most valuable from among the available products.
The e-commerce environment provides a number of interesting challenges to the RS developer. One major challenge is in the area of so-called hybrid systems. The currently dominant systems generate recommendations utilizing only one type of input data about customer preferences for products (for example, explicit ratings data or purchase data). The goal of hybrid system is to take all available preference data simultaneously, and use it in an intelligent way to provide recommendations.
RSs are a very popular and effective paradigm in retail business. With RS shoppers can find items they like with less effort. Furthermore, they are presented with items they've never thought of buying, but which actually suit their needs.
RSs are useful when users have not rated most of the products. If different groups of users rate different categories of products, it becomes less likely than rated products and can be used to generate recommendations.
Although dimensionality reduction algorithms are used to solve this problem, they are ill-suited to extremely scarce data and should be modified for RSs. While a large amount of data will slow down the system, a lack of data will also hamper the ability to generate recommendations.

In order to overcome this problem, we present a RS based on content-based filtering and machine learning (ML) which is defined as an artificial intelligence technology that allows machines to learn without being specifically programmed for this purpose. Machine leaning is generally classified into three broad categories algorithms: supervised, unsupervised, and reinforcement learning.in our system we used the two algorithm KNN and linear kernel which are classify in supervised algorithm type , these algorithms are composed of three categories are ;decision trees that is based on a graph model ;association rules can be defined as a truth table that results from the combination of two or more characteristics ;and Clustering consists in the division of the population or data point ,for example the KNN ,K-means algorithms.

IV.2 Proposed Methodology

Our RS is based on the content-based filtering and uses a combination of two ML algorithms to better generate product recommendations.
Our system is able to analyze products and similarity with the active user for products recommendation. The proposed approach also addresses the problems of the previous methods to efficiently evaluate user preference on products and balance feature analysis. The K-NN classification method has been qualified online and real-time to find user behavior data and containing the relationship between the similarity user behaviors to get an idea of the preferred products.and it will classify the features using race classifiers from both quality and quantity aspects.
This methodology provides recommendations based on what the user has reviewed in the past. The content is described using labels and the labels are weighted according to the quality of their product description. By applying the algorithm linear kernel to represent the similarity score of each product category with all other product categories in the matrix, and then using the KNN with these labels and the user's preferences to have the neighboring products of these categories giving at the end a recommendation of other products to the active user.
The proposal is an automatic approach and mainly involves six steps. The first step "Data collection" where we import our database, the second step "Data cleaning" this step consist of cleaning data which refers to the correction of error for example verification of duplicate word ,Tf-IDf , the next step "Analyse user Behavior"this step based to create the user's profile , then we moved to"Formation of the Vector of Features to Each Neighbor and Products with KNN " in this step we decompose oud data product category in different sub category ,and "Classification with linear Kernel" using a linear classifier to solve a nonlinear problem ,And finally, the last step "Application of a UDF function"where we takes the name of the product as input and produces the most similar product.

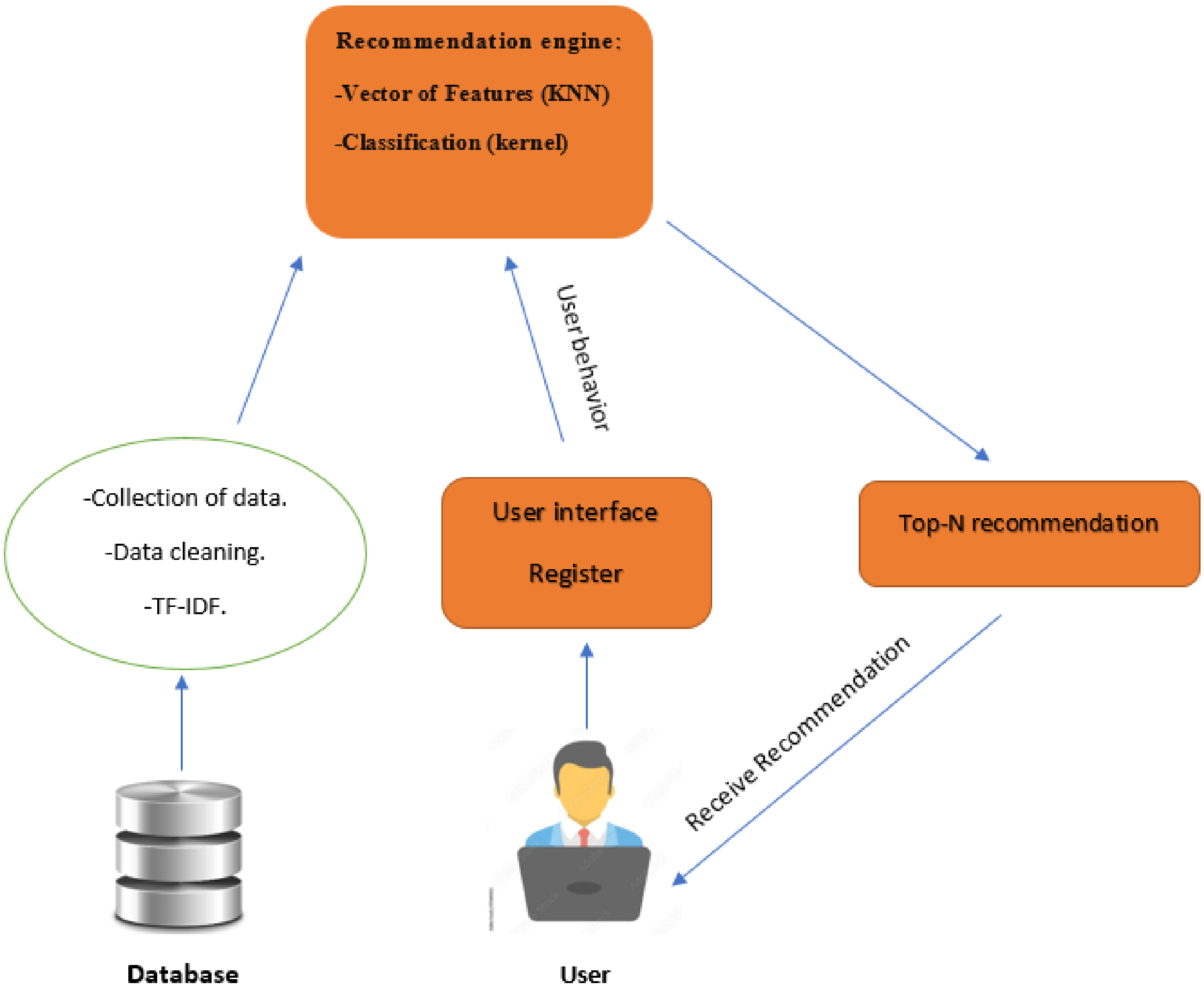


Figure IV.1: System architecture

IV.2.1 Data Collection:

The first step of the proposal is to import the extracted dataset from Amazon and define it with their information. Therefore, this part is very important to start the referral process and use it to build connected reports to improve daily service or test acceptance of a new offer.

The dataset is collected from Amazon in csv format, this database contains (product name, category, price, characteristics, ...). In this table, we present some field of our database and their description.

Database	Type	Description
Unique id	Integer	Is the identifier of each product
Product name	String	Is the name of the product
Category	String	There are many categories so we decompose it into three of sub categories
Selling price	Float	Is the price of each product
Product url	Varchar	Is the link that we can find the product we want

Table IV.1: database description

IV.2.2 Data cleaning

In the next step, we have to cleaning data which refers to the correction, the correction or delete incorrect, corrupted, improperly formatted, duplicate, or incomplete data in a dataset. To have a quality dataset, we first check the characteristics of the variables, then we eliminate all duplicate words, remove redundant columns, remove zero values to avoid errors, filtering and weighting of TF words -IDF. Here is a detailed description of the data set cleansing steps:

- **Verification of the characteristics of the variables**
  This characteristic makes it possible to define a different type of variable, for example:

```
# Column Non-Null Count Dtype
--- ------ -------------- -----
 0 Uniq Id 10002 non-zero object
 1 Product name 10002 non-zero object
 2 Brand name 0 non-zero float 64
```

- **Verification of duplicate words**
  Another feature, is a technique used to reduce the amount of storage space an organization needs to back up its data in order to we eliminate all duplicate data words, For example :

```
10,000 False
10001 False
Length: 10002, type: bool
All 10002 lines are unique.
```

- **Removal of redundant columns**

  In this case, we delete all the columns of the same name in addition to the first occurrence of the columns.

In this example we have 7216 non-zero Product Name and the same is true for the category.

```
# Type of number of columns not zero
--- ------ ------------- -----
 0 Product name 7216 non-zero object
 1 non-zero object of category 7216
```

- **Filtering**

  Filtering the data in a data frame is one of the most common operations when cleaning data. Pandas offers a wide variety of methods for selecting data based on the position and label of rows and columns. In addition, Pandas also allows you to get a subset of data based on column types and filter rows with Boolean indexing.

  In our system, we divide our data categories into three subcategories, the first category called main category, the second named subcategory, the third called secondary category and the last category consists of the remaining categories called other category.

- **TF-IDF word weighting**

  TF-IDF stands for "Term Frequency - Inverse Document Frequency". Is a statistical measure that assesses the relevance of a word to a document in a collection of documents. We generally calculate a weight for each word which signifies the importance of the word in the document and the corpus. This method is a technique widely used in information retrieval and text mining.

  Frequency of terms (TF)is a statistical measure that assesses the relevance of a word to a document in a collection of documents.Term frequency is commonly used in text mining, machine learning, and information finding tasks. Because documents may have different lengths, a term may appear more frequently in longer documents than in shorter documents. **Reverse Document Frequency (IDF)**is the inverse of the frequency of the document which measures the informativeness of the term t. When we calculate IDF, it will be very low for the most frequent words such as stop words. The time it takes to complete a cycle is the period. Frequency is the reciprocal of this, the number of cycles per second. The distance that sound travels during a period is the wavelength. TF-IDF is the simple product of TF and IDF, so that both drawbacks are solved, by making predictions and looking for relative information. Each word or phrase has its score for TF and IDF. In addition, the result of the TF and IDF product of a term refers to the TF-IDF weight of that term. Thus, we can say that the rarer the term, the higher the TF-IDF (weight) score. Therefore, the TF of a word is the frequency of a word, while the TF-IDF of a word is a measure of the importance of that term in the corpus. If the words have a high TF-IDF weight, their content will always appear among the first search results. So we can stop bothering ourselves with stop words. In addition to finding words with higher search amount and lower competition with success. The formula of the TF-IDF combines that of the two indicators that compose it.

Wx, y = TFx, y * log (N / DFx)or :

- Wx, y denotes the relevance score of any term x in a document y (TF-IDF);
- TFx, y denotes the frequency of the term x in the document y;
- DFx denotes the number of documents containing x;
- N is the total number of documents.

IV.2.3 Analyse user Behavior:

This step is to create the user's profile based on the information with different methods. After registration, the user behavior form is introduced for each user.

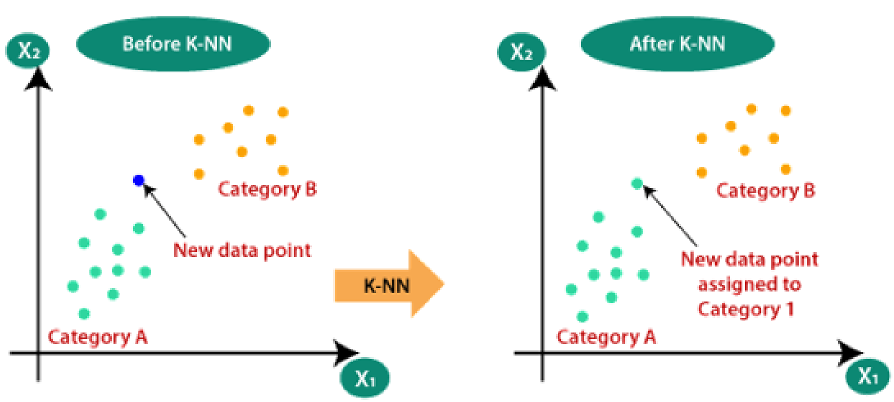
IV.2.4 Formation of the Vector of Features to Each Neighbor and Products with KNN

In this step we use the KNN who is a supervised machine learning algorithm. The algorithm can be used to solve both the classification and regression problem. The data is assigned to the class that has the nearest neighbors. The KNN algorithm works by different steps, we calculate the distance between the test data and each row of training data., we sort the

calculated distances in ascending order based on the distance values. Then, we Get the first k rows of the sorted array, we get the most frequent class of these rows, finally return the predicted class [12].
In our database example, we have decompose the product categories into 3 subcategories after we apply the KNN. Suppose there are two categories. Category A and Category B, and we have a new data point x1, so that data point will fall into which of those categories. To solve this type of problem, we need a KNN algorithm. With the help of KNN, we can easily identify the category or class of a particular dataset.
In this, the active user has purchased products from the list of categories like several products in electronics, dresses and so on. When the closest Neighbors are defined, it is important to see if the ID of the products acquired is new to the active user. Although the Products identify with neighbors, for every product, The framework will prescribe the number of products with the most minimal Euclidean distance esteems to the user.
The Euclidean distance (D) between the two data points x1 and y1 as,[12] where x1 and x2

$$D(x1, x2) = \sqrt{(\sum_{S=1}^{M}(x1 - x2)^2}$$

denote the input and feature instances.
Consider the diagram below: [13]



IV.2.5 Classification with linear Kernel

In machine learning, a "kernel" is typically used to refer to the kernel trick, a method of using a linear classifier to solve a nonlinear problem. The kernel function is what is applied on each data instance to map the original nonlinear observations into a higher dimensional space where they become separable. Used when the data is linearly separable, that is, it can be separated using a single row. It is one of the most commonly used kernels. It is mainly used when there is a large number of entities in a particular dataset.
In our system, we use a matrix to represent the similarity score of each product category with all other product categories using the kernel algorithm [14] [15].

```
[0. 1.27876877 ... 0.25244281 0.27876877 0.]
[0. 1.27876877 ... 0.25244281 0.27876877 0.]
[0.76159416 0.76196458 0.76169747 ... 0.76168771 0.76169747 0.76159416]
```

IV.2.6 Application of a UDF function

The User Defined Function (UDF) is a function that we define ourselves to perform certain specific tasks are called user-defined functions. In this code, we have defined our function which takes the name of the product as input and produces the most similar product [16]. It's returned the top-N recommendation for each user from a set of predictions.

```
Args:
    predictions(list of Prediction objects): The list of predictions, as
        returned by the test method of an algorithm.
    n(int): The number of recommendation to output for each user. Default
        is 10.

Returns:
A dict where keys are user (raw) ids and values are lists of tuples:
    [(raw item id, rating estimation), ...] of size n.
'''
```

we present the steps of our approach in the form of this algorithm

Algorithm : Algorithm for Product Recommendation

Input: User, U appartient u1, u2,... .ui user behavior matrix, product set.

P = (p1, p2,... .pi)

Output: Recommendation list for Utarget.

1. Dt ← Database is loaded.
2. Uk ← Acquire user behavior data from the web access log database
3. User behavior matrix (WK) ← each user ui
4. For (Each User ← U)
5. User behavior form (Uk) ← Ui * Ui! Number of users *
6. Sort similarities Usersim (Pi, Utargetn) * Utargetn! Number of target user *
7. }
8. Nnon ← zero ← Number of non-zero rows in the ranking list.
9. If (WK ? Is) {* Is! Input Sequence *
10. Return Ranking List
11. }
12. For (User U: Start the iteration for similarity) {
13. Ru ← Recommendation Engine has user behavior features (Pn, Vn, CA, Pn).
14. Nf ← Normalize the formation of the vector of features.
15. C (x) ← Classify class label features for Pn
16. f (x, y) ← Divides the function for each user ui
17. C (x) ← Weighted K-NN (C (x)) to employ the similarity feature
18. If (sim (Utarget, ui) _P){
19. Return (recommended Product List)
20. Sort List (recommended Product List)
21. Set of products recommended to the Utarget
22. }

Conclusion

In this chapter, we have presented our approach for building a RS based on content filtering and ML algorithms. We have defined the system architecture and the different steps of the proposal in detailed.

In the next chapter, we will discuss the experimentation and evaluation of the proposal.

Chapter V

Experimentation

Introduction

Recommender systems are being used by an ever-increasing number of E-comme- rce sites to help consumers find products to purchase. What started as a novelty has turned into a serious business tool. Recommender systems use product knowledge either hand-coded knowledge provided by experts or "mined" knowledge learned from the behavior of consumers to guide consumers through the often-overwhelming task of locating products they will like. In this article we present an explanation of how recommender systems are related to some traditional database analysis techniques. We examine how recommender systems help E-commerce sites increase sales and analyze the recommender systems . Based on these examples, we create a taxonomy of recommender systems, including the inputs required from the consumers, the additional knowledge required from the database, the ways the recommendations are presented to consumers, the technologies used to create the recommendations, and the level of personalization of the recommendations.

In this chapter, we will describe the navigation diagram between the interface and our application, as well as the different hardware and software environments used in the development of our system. Then we will evaluate our system and test its effectiveness. Finally, we will develop and explain the different interfaces of our application.

V.1 Design and specification

Our system consists of an authentication page and a registration page (Form to be completed by the user) and a home page visible to all users, each product contain a category; a selling price; description of a product; a specification and shipping weight (pounds). And, each category we have divide it in three-sub category named main, sub and side category. Finally, in order to show how the different algorithms are implemented, we have added an analysis page that will show statistics and graphics of our system. Our application decomposes into three phases:

- The first phase, determine page authentication that allows user to connect with unique username and password to access the application.
- The second phase, the user asked for a product recommendation then this application research in her database, and return the entire similar product to the product or the re-search word we asked.
- The third one, consist to the choice we have recommended, is defined by the provision of certain information necessary for the smooth running of the recommendation.

First, when we go to access to our application, first all we have home page that is the first interface of our application. It is contain of ‘searches bar ‘which is a graphical interface component which allows you to enter a character string to search for a product that the user want to. And we have in the right space three bottoms, the first is ‘ Log in ‘ and the second ‘ Card ’ and the third is ’For Me’.

- ‘Log in’ that is the connection field where the user authenticate, if the user has an ac-count he will login, if not he will sign up with username and a password. If the user is connected, the bottom ‘log in’ will be ‘My account’ bottom where we found our products and we can logout in the same interface ‘My account’.
- ‘Card’ this bottom allows user to enter to a new interface which contain a table of product that we recommended ,each product contain a price and quantity we want to added in our card, which have a summary where found a total invoice.
- ’For Me’ this bottom allows user to found all his similar product to his research recommendation, it makes your research easy.
- In the same interface we found too a space where we have the latest product.

V.2 Experimentation environment

V.2.1 Hardware environment

All the experiments were carried out with an Intel Core i5 CPU with a frequency of 3.1 GHz and 8 GB of memory running under the Windows 10 platform.

V.2.2 Software environment

We have implemented our web application using the Django framework with HTML5, CSS3, and Bootstrap. The learning algorithms used to calculate the recommendations were coded in python; we used the CSV (Comma-Separated Values) for the Amazon database.

Python

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics developed by Guido van Rossum. It was originally released in 1991. Python has a reputation as a beginner-friendly language, replacing Java as the most widely used introductory language because it handles much of the complexity for the user, allowing beginners to focus on fully grasping programming concepts rather than minute details [1].

- **Python Use Cases** Creating web applications on server

  Building workflows that can be used in conjunction with software

  connecting to database systems

  Reading and modifying files

  performing complex mathematics

  Processing big data

  Fast prototyping

  Developing production-ready software [1].

  Here are some libraries used in our implementation:

  - **Scikit-learn:** is probably the most useful library for machine learning in Python. The sklearn library contains many efficient tools for machine learning and statistical modeling including classification, regression, and clustering and dimensionality reduction.

  - **Pandas:**Is a Python library for data analysis, It presents a diverse range of utilities, ranging from parsing multiple file formats to converting an entire data table into a NumPy matrix array. This makes pandas a trusted ally in data science and machine learning.

Django

Is an open-source python web framework used for rapid development, pragmatic, maintainable, clean design, and secure websites, Django is an open-source python web framework used for rapid development, pragmatic, maintainable, clean design, and secure websites. Django is based on MVT (Model-View-Template) architecture. MVT is a software design pattern for developing a web application. MVT Structure has the following three parts – Model:

- **Model:** The model is going to act as the interface of your data. It is responsible for maintaining data. It is the logical data structure behind the entire application and is represented by a database (generally relational databases such as MySQL, Postgres).
- **View:** The View is the user interface — what you see in your browser when you render a website. It is represented by HTML/CSS/JavaScript and Jinja files.
- **Template:**A template consists of static parts of the desired HTML output as well as some special syntax describing how dynamic content will be inserted.
  See the following graph that shows the MVT based control flow ,Here, a user requests for a resource to the Django, Django works as a controller and check to the available resource in URL.If URL maps, a view is called that interact with model and template, it renders a tem-plate. Django responds back to the user and sends a template as a response [2].

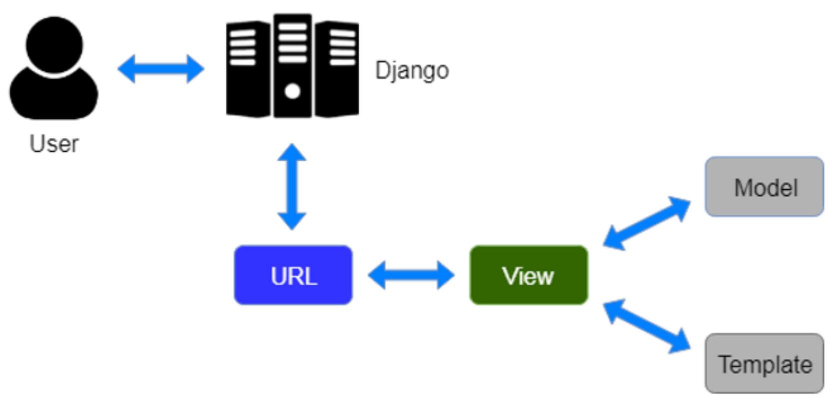


Figure V.1: Mvt architecture

Anaconda Navigator

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda® distribution that allows you to launch applications and easily manage conda packages, environments, and channels without using command-line commands. Navigator can search for packages on Anaconda.org or in a local Anaconda Repository. It is available for Windows, MacOs, and Linux [3].

What applications can I access using Navigator?

The following applications are available by default in Navigator [3]:

- JupyterLab
- Jupyter Notebook
- Spyder
- PyCharm
- VSCode
- Glueviz
- Orange 3 App
- RStudio
- Anaconda Prompt (Windows only)
- Anaconda PowerShell (Windows only)
- Anaconda Prompt (Windows only)
- Anaconda PowerShell (Windows only)

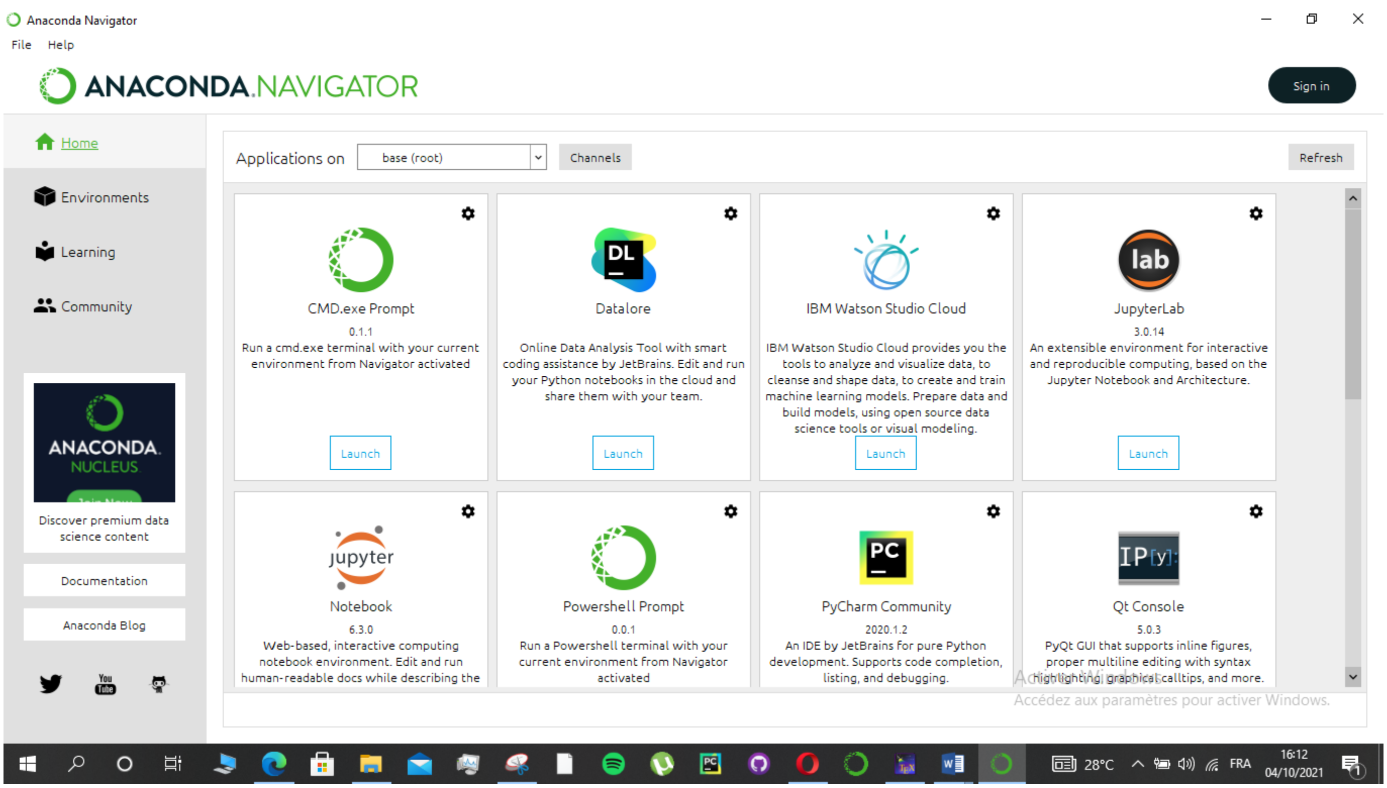


Figure V.2: Anaconda Navigator

- **Prompt Anaconda:**

  Is a library for prompting input on the command line for Python 3.3+. It is pure Python code with no dependencies.

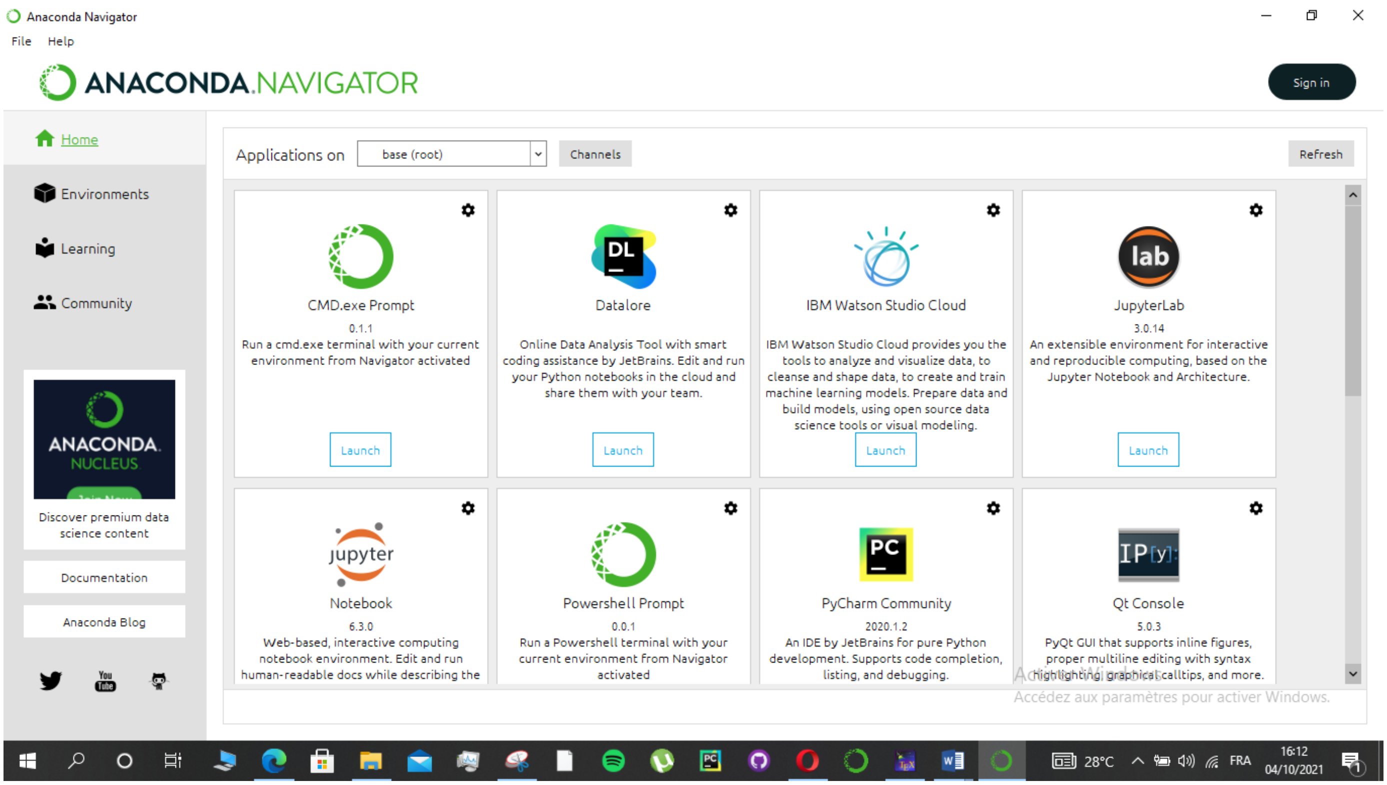


Figure V.3: Anaconda Environment

- To install and create environment we use :

  Conda create – name myenv python 3.6

  Proceed ([y]/n)? y

- Enter and exit environment :

  Enter: (base) c: users vmi - activate myenv, exit: (myenv) c: users vmi - desactivate

- Install tensorflow: copy the download link.

- Import tensorflow :

  (myenv) c: uses vmi - python

  - import tensorflow

Node js

Is an open source development platform for executing JavaScript code server-side. Node is useful for developing applications that require a persistent connection from the browser to the server and is often used for real-time applications such as chat, news feeds and web push notifications [4].

- Concepts :

  The following diagram depicts some important parts of Node.js, which we will discuss in detail in the subsequent chapters.

Where to Use Node.js?

Following are the areas where Node.js is proving itself as a perfect technology partner [4].

- I/O bound Applications
- Data Streaming Applications
- Data Intensive Real-time Applications (DIRT)
- JSON APIs based Applications
- Single Page Applications

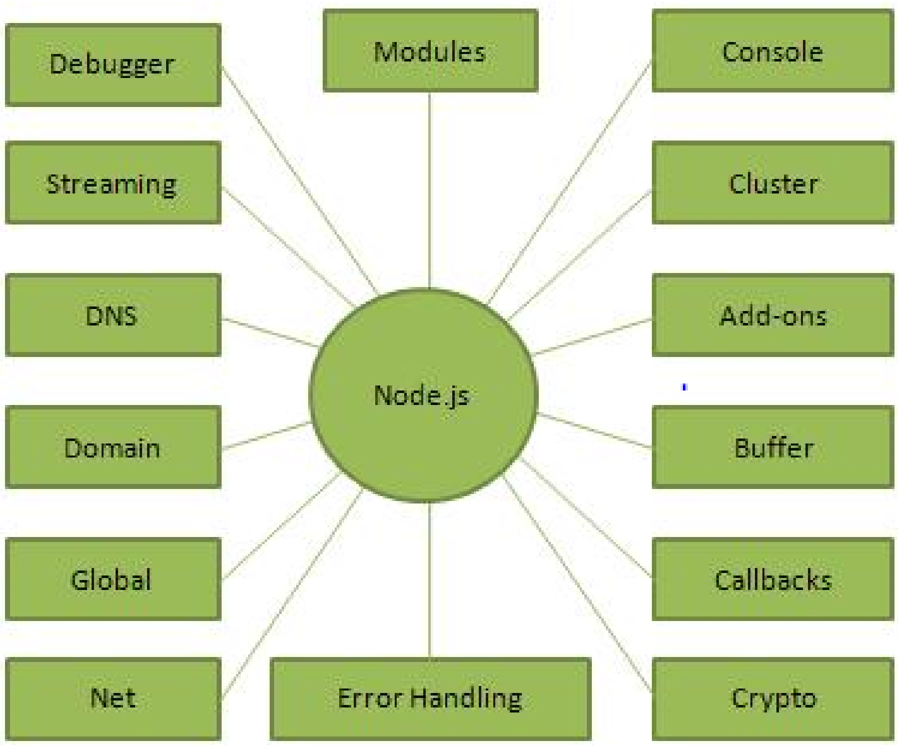


Figure V.4:

CSV file

A CSV (comma-separated values) file is a simple text file in which information is separated by commas, and is common in spreadsheet apps. CSV files are most commonly encountered in spreadsheets and databases. You can use a CSV file to move data between programs that are not ordinarily able to exchange data [5].

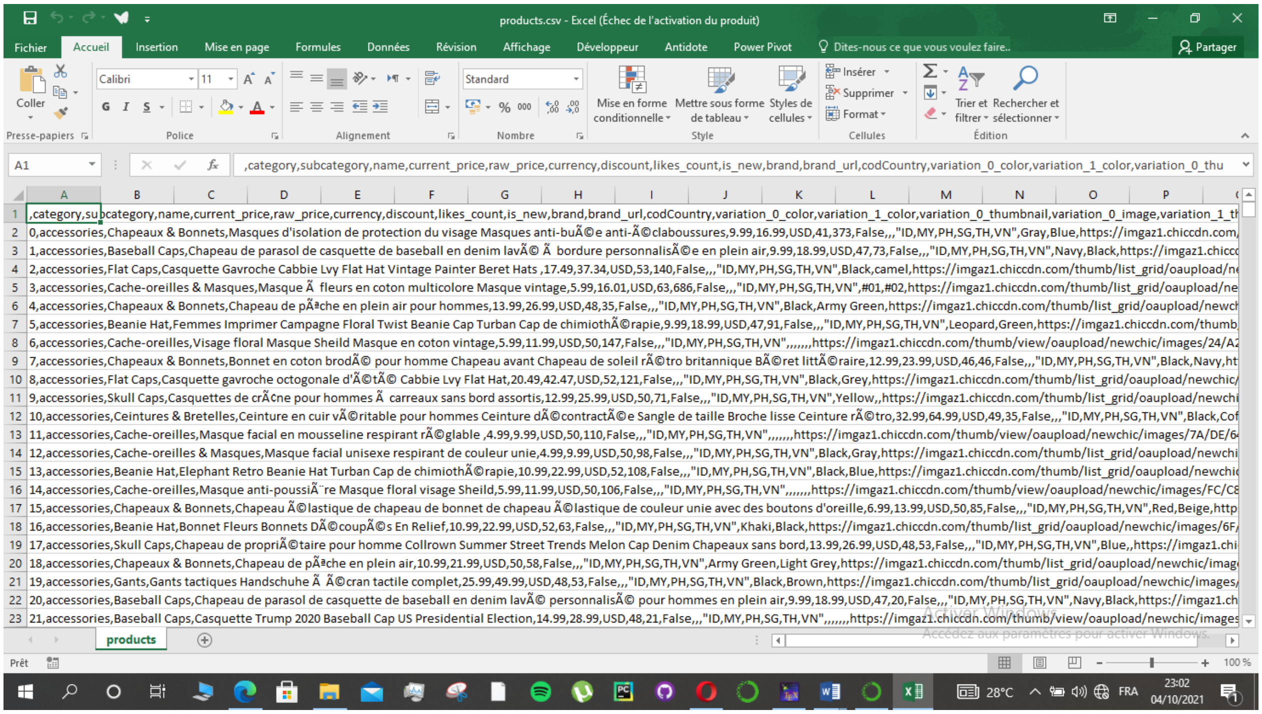


Figure V.5: CSV data form

DB browser for SQLite

DB Browser for SQLite is a visual tool used to create, design, and edit database files compatible with SQLite. The current iteration is a successor to what was formerly known as SQLite Database Browser. DB Browser is a free, open-source tool for end users and developers who want to use a familiar spreadsheet-like interface to manage database files without having to learn complicated SQL commands [6].

- **Import and export file:**

DB Browser for SQLite has a suite of import and export tools that allow users to import and export the following:

- Records as text
- Tables from/to CSV files
- Databases from/to SQL dump files

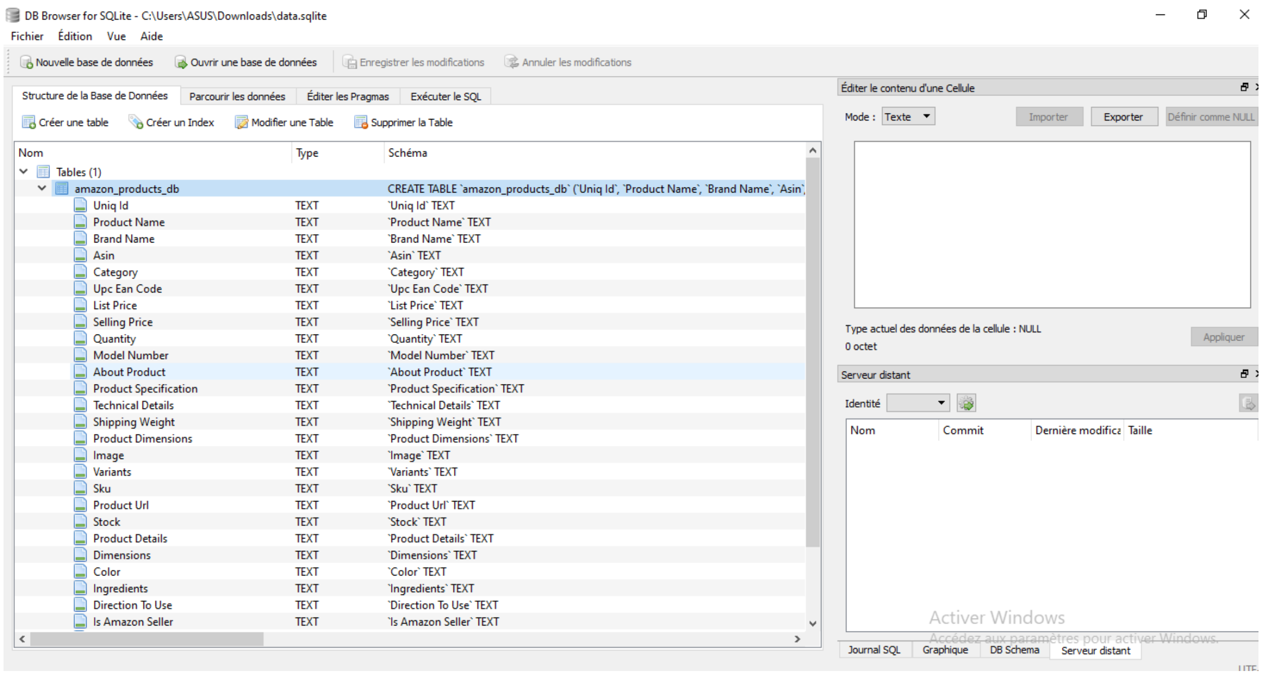


Figure V.6: SQLite data

V.2.3 Dataset

In our system, we have used a dataset downloaded from Amazon, which contain a lot of product that made up of several fields ,categories, product name ,Selling price, shipping weight about this product, product specification and product url.

V.3 Implementation

This section describe the different part implemented of the recommendation system, and explain each part of our application. We will also present different interface, Home page, login page, card page.

V.3.1 Home page

This page allows you to view the first interface of our application, and a space where we found our latest products. We have too a search bar where we can do our research product.

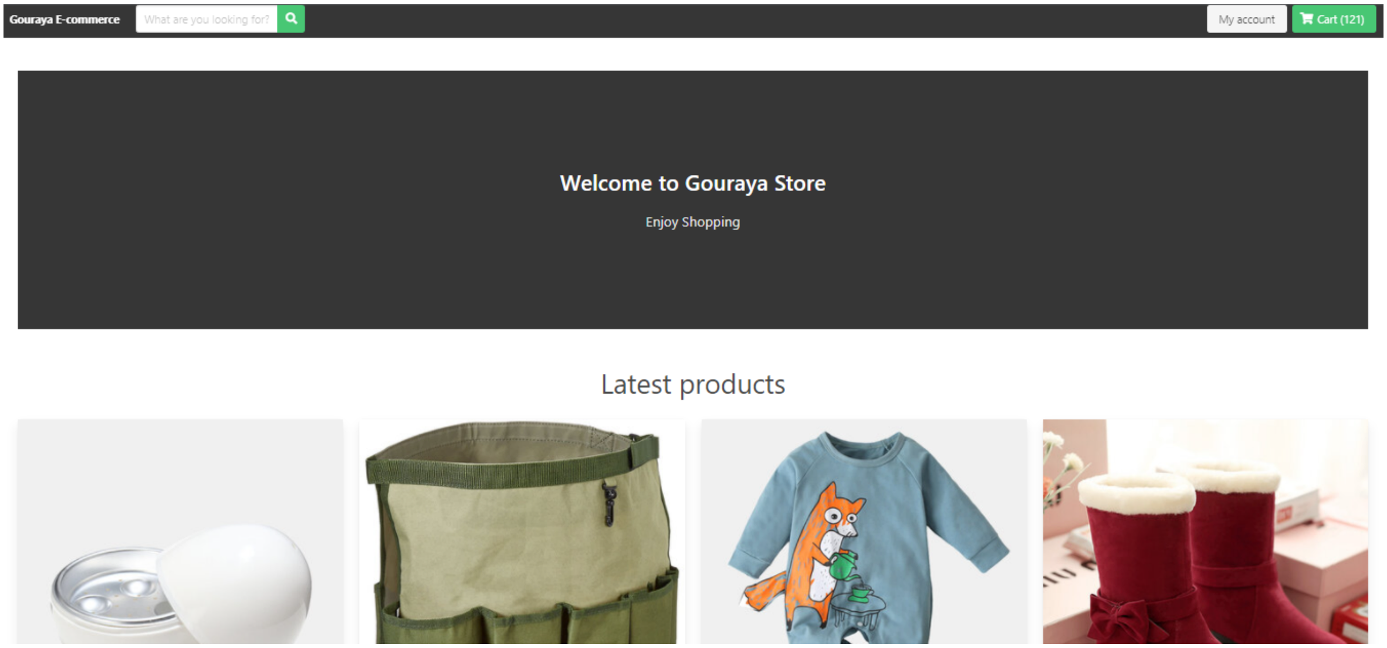


Figure V.7: Home page

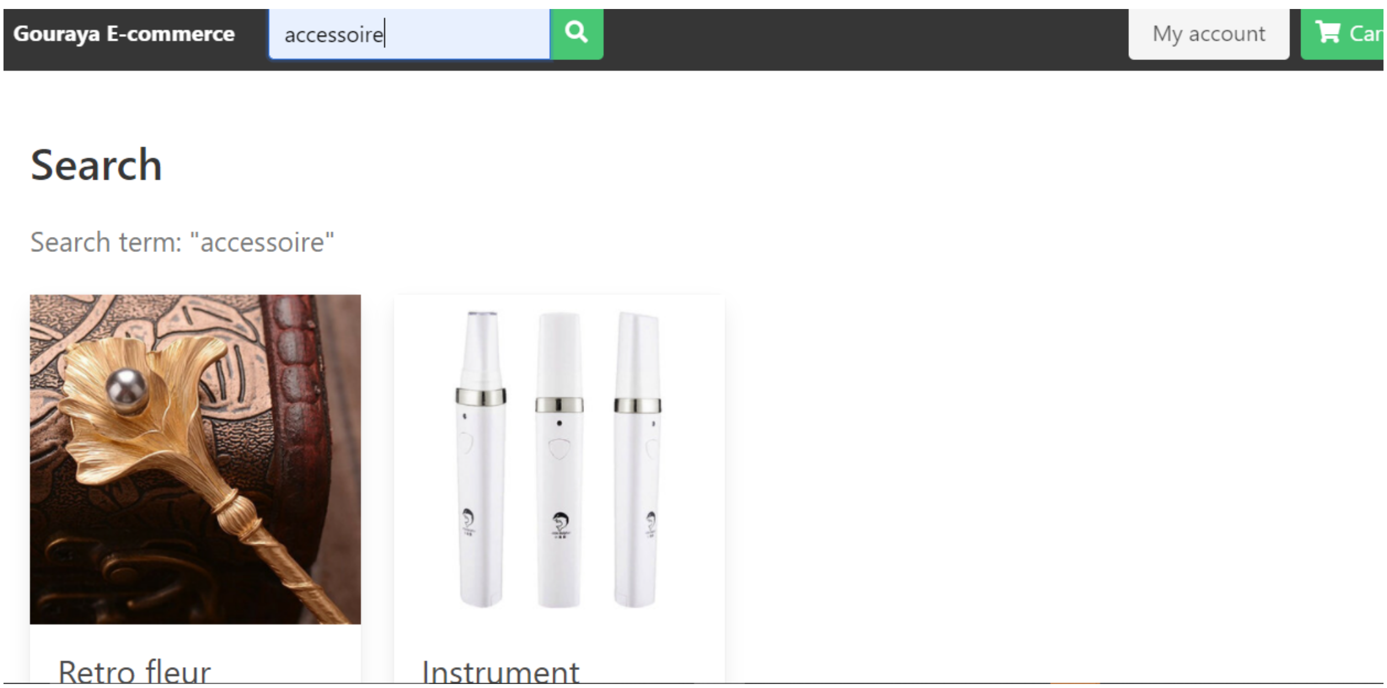


Figure V.8: Search product

V.3.2 Login page

In this page the user can authenticate to the application by using his username and password to access in the next interface as we see in the figure (IV.9), if we have an account then we login figure(IV.10:).in the same interface we can logout figure (IV.11).

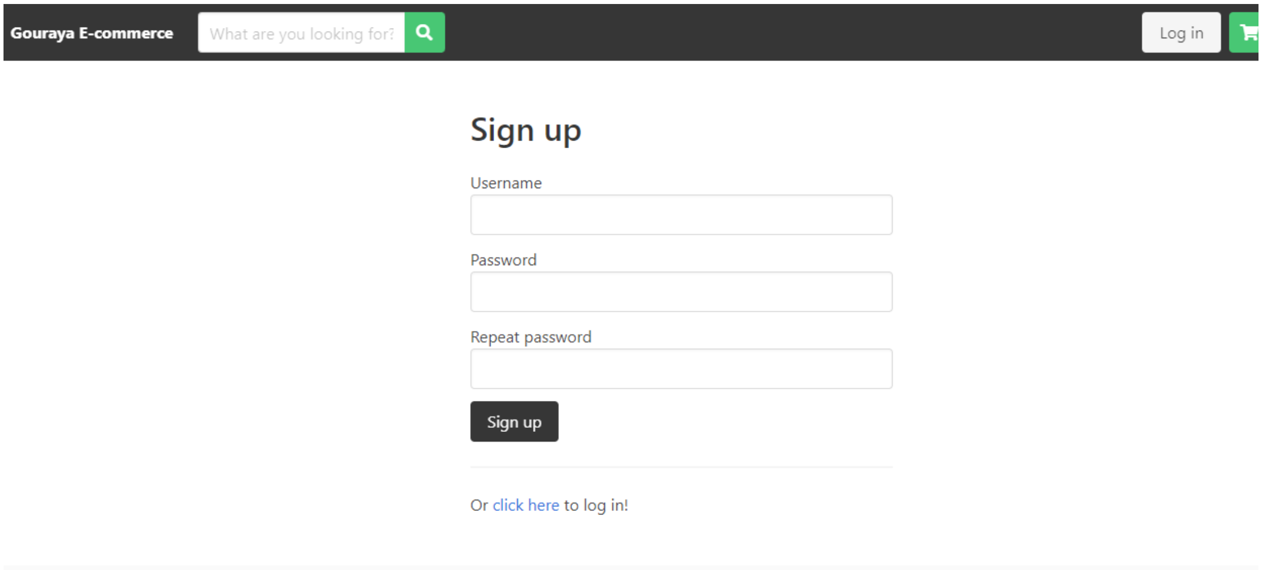


Figure V.9: Sign Up page

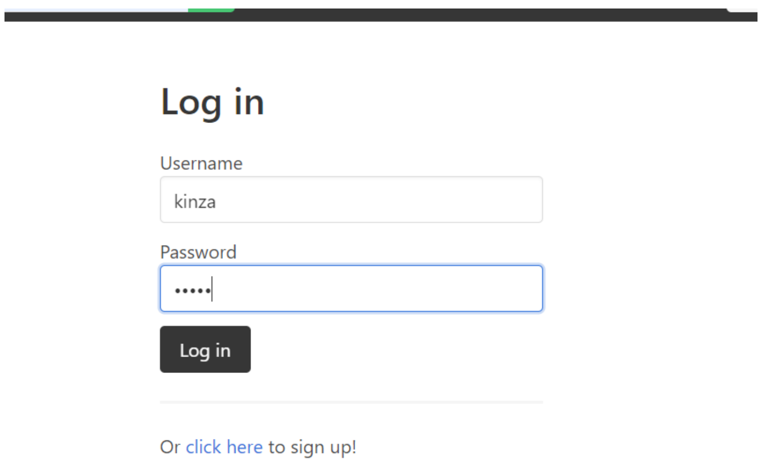


Figure V.10: Login page

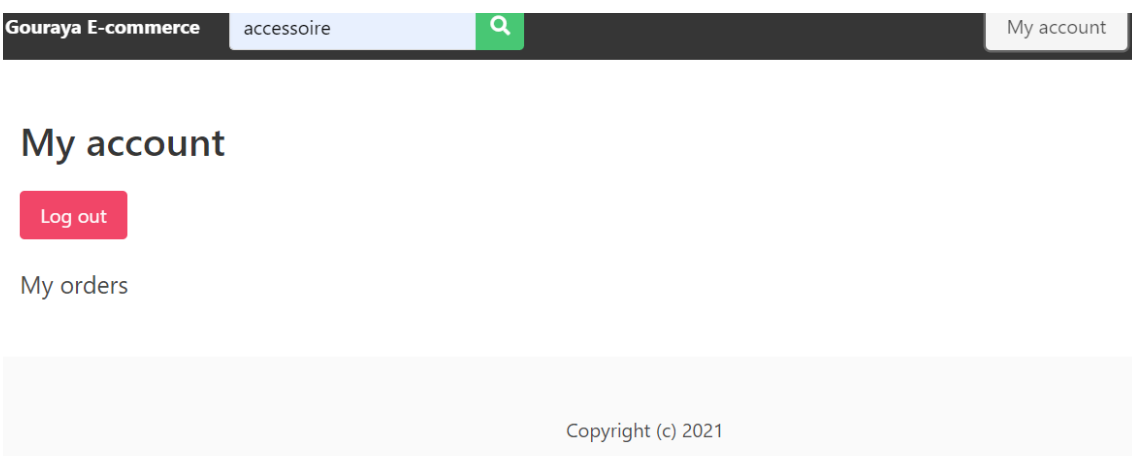


Figure V.11: Logout page

V.3.3 Recommended product

In this page we found all the product similar to the product that we recommended.

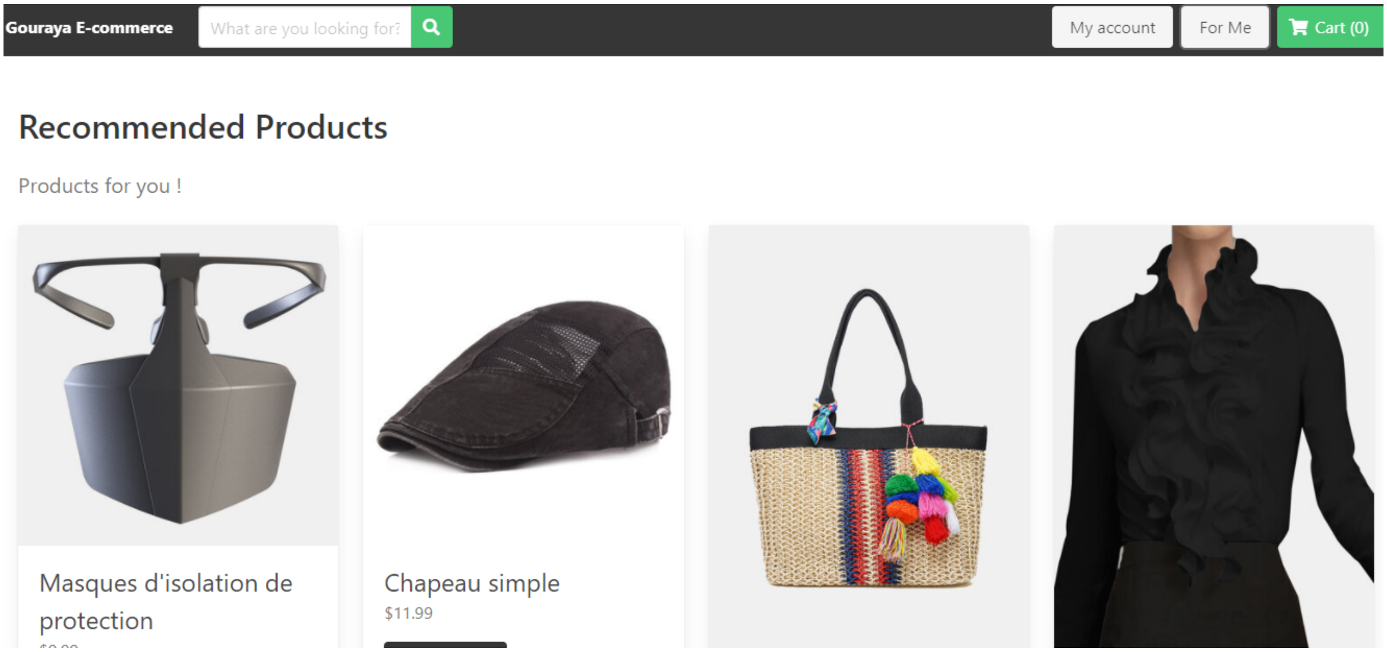


Figure V.12: For Me

V.3.4 Card page

This page allows users to observe all the product that he recommended.

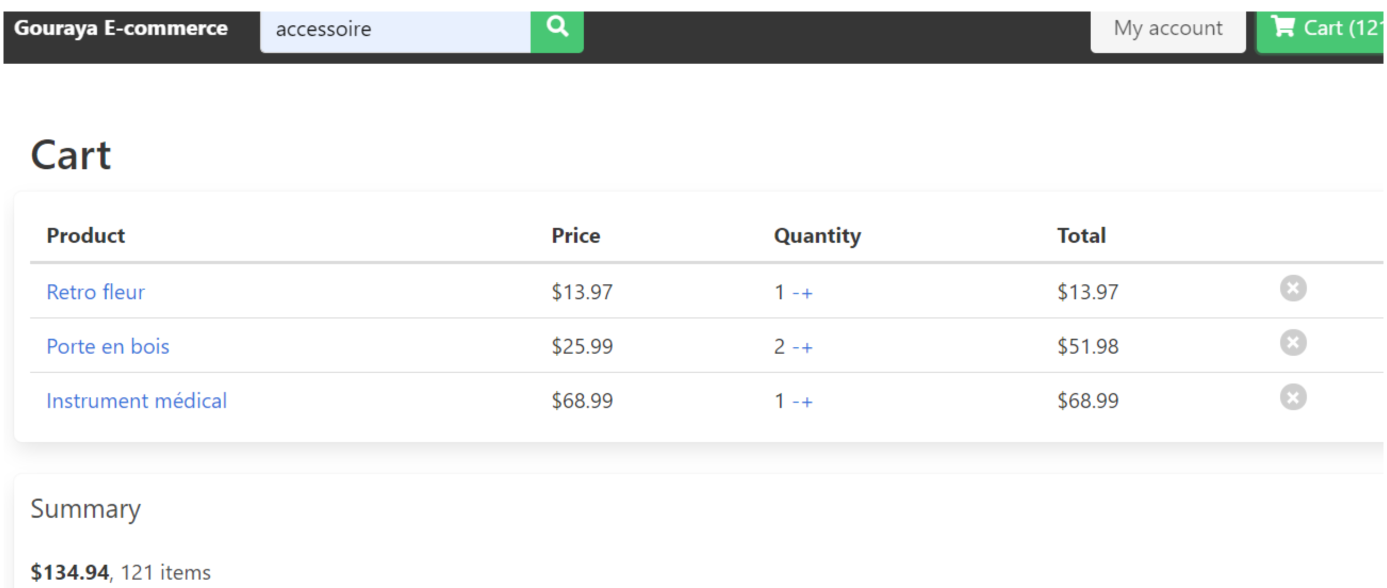


Figure V.13: Card page

V.3.5 Evaluation

The quality of a recommendation algorithm can be evaluated using different types of measurement, which can be accuracy or coverage. The type of metrics used depends on the type of filtering technique. Accuracy is the fraction of correct recommendations out of total possible recommendations while coverage measures the fraction of objects in the search space the system is able to provide recommendations for. Metrics for measuring the accuracy of recommendation filtering systems are divided into statistical and decision support accuracy metrics. The suitability of each metric depends on the features of the dataset and the type of tasks that the recommender system will do.
Statistical accuracy metrics evaluate accuracy of a filtering technique by comparing the predicted ratings directly with the actual user rating. Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Correlation are usually used as statistical accuracy metrics. MAE is the most popular and commonly used; it is a measure of deviation of recommendation from user's specific value. It is computed as follows (1):

$$\text{MAE} = \frac{1}{N} \sum_{u,i} |P_{u,i} - r_{u,i}| \text{ .... } (1)$$

Where P(u, i) the predicted rating for user u on item i is, r(u, i) is the actual rating and N is the total number of ratings on the item set. The lower the MAE, the more accurately the recommendation engine predicts user ratings. Also, the Root Mean Square Error (RMSE) is given as :

$$\text{RMSE}= \sqrt{\frac{1}{n}} (P_{u,i} - r_{u,i})^2 \text{.................... } (2)$$

Root Mean Square Error (RMSE) in equation (2) puts more emphasis on larger absolute error and the lower the RMSE is, the better the recommendation accuracy. Decision support accuracy metrics that are popularly used are Reversal rate, weighted errors, Receiver Operating Characteristics (ROC) and Precision Recall Curve (PRC), Precision, Recall and F-measure. These metrics help users in selecting items that are of very high quality out of the available set of items. The metrics view prediction procedure as a binary operation, which distinguishes good items from those items that are not good. ROC curves are very successful when performing comprehensive assessments of the performance of some specific algorithms. Precision is the fraction of recommended items that is actually relevant to the user, while recall can be defined as the fraction of relevant items that are also part of the set of recommended items. They are computed as in (3) and (4) :

$$\text{Pression} = \frac{\text{Correctly recommended items}}{\text{Total recommended items}} \;\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

$$\text{Recall} = \frac{\text{Correctly recommended items}}{\text{Total useful recommended items}} \;\ldots\ldots\ldots\ldots\ldots\ldots\; (4)$$

F-measure defined equation (5) below helps to simplify precision and recall into a single metric. The resulting value makes comparison between algorithms and across data sets very simple and straightforward.

$$\text{F-measure} = \frac{2PR}{P+R} \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\; (5)$$

Coverage has to do with the percentage of items and users that a recommender system can provide predictions. Prediction may be practically impossible to make if no users or few users rated an item. Coverage can be reduced by defining small neighborhood sizes for user or items.

Conclusion

In this chapter, we have introduced the design and specification of our application. Then we have moved to the both of environment, the hardware and the software where we have mentioned all the software we used in our system. Then we presents all the interface steps of the site in the implementation part, finally we gave the evaluation of our work.

Chapter VI

General conclusion

In the Internet, where the number of choices is overwhelming, there is a need to classify, filter and efficiently deliver applicable information to alleviate the problem of information overload, which has generated a potential problem for many Internet customers.

RS solves this problem by exploring a large volume of dynamically generated information to provide customers with personalized content and services. The plurality of RF approaches and algorithms can be used for various types of recommendations depending on the domain for which they are submitted. The choice of the specific technique to be used in an RS depends on the results of the required recommendation. Therefore, improvements in a recommender technique must be supported by the corresponding evaluation measure. Most modern e-commerce systems are geared towards adding recommendation systems as they increase the efficiency of the shopping site by displaying the most popular products.

Project Reminder

In this thesis; it is shown that an e-commerce RS based on the proposed algorithm can provide consumers with the most needed information at the most appropriate time and offer them a more comfortable shopping experience. The Current literature concerns RSs that collect information only on the purchasing platform. According to the characteristics of the e-commerce recommendation system, the proposed system collected consumer requirements from several approaches. It can provide consumers with more specific recommendations. Accurate recommendations can not only improve customer satisfaction, but also affect the conversion rate of merchandise. Thus, it is advisable for e-merchants to recommend with precision the most suitable products in order to boost sales. Electronic trailers must take advantage of a user's multi-source information to guide and cultivate consumer decision-making by combining the user's mobile shopping habits. They can implement specific recommendations through research on user demand, consumer behavior and product category preferences.

Thus, this research enriches the theory of product prices so that a consumer or a customer can have his choice with a beating price to recommend him products by going from the cheaper price to an expensive price. At the same time, our brief is of great importance to e-commerce platform providers in the design of a product RS. E-commerce platform operators could redesign the algorithms of their RS based on the above in order to choose more suitable recommended products. For that we used the Django platform to better present the system trousers. The referral system could affect the privacy of consumers and suppliers. And to avoid customer and supplier information security problems, we have chosen that our system must have an authentication part. On the other hand, in the process of designing a RS, we only focus on the convenience of a user as to how to precisely find the products they are interested in.

Perspectives and future work

- In terms of speed, the proposed algorithm has no advantage but it is not really slow, in the next work we will look at how to improve the speed of the system.
- Apply another evolutional approach of the RS in order to have better results.
- Provide point of delivery customization: The RS directly customizes the point of delivery for the E-commerce site.
- Provide quick response throughout the value chain: We predict that RS will be used in the future to predict demand for products, enabling earlier communication back the supply chain.
- In future research, we want to develop a useful tool to target consumers with personalized messages and put multi-source information fusion technology to improve the prediction accuracy of personalized recommendations in the e-commerce terminal on mobile thereof. discusses the specifics of a mobile platform compared to a personal computer (PC).
- And finally, design a more efficient and accurate RS for meeting the needs of the users.

Bibliography

[1] Idir Benouaret. Un système de recommandation contextuel et composite pour la visite personnalisée de sites culturels.Autre [cs.OH]. Université de Technologie de Compiègne,2017.

[2] Malone, T., Brobst, S., Cohen, S., Grant, K., and Turbak, F. (1987). Intelligent information des systemes de partage. In Communications of the ACM, volume 30, pages 390–402.

[3] Resnisck, P. and Varian, H. (1997). Recommender systems. In Communications of the ACM, volume 40, pages 56–58.

[4] Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction, 12(4) :331–370.

[5] A. T. NGUYEN. COCoFil2 : Un nouveau système de filtrage collaboratif basé sur le modèle des espaces de communautés. Thèse. Université joseph Fourier Grenoble I. Novembre 2006.

[6] Maes, P. and Shardanand, U. (1995). Social information filtering: algorithms for automating "word of mouth". In the SIGCHI conference on Human factors in computing systems, Denver, Colorado, United States. ACM Press/Addison-Wesley Publishing Co.

[7] `https://www.ionos.fr/digitalguide/web-marketing/vendre-sur-internet/les-systemes-de-recommandation-pour-le-e-commerce/`

[8] Sachin Gulabrao W.,Kishor S., An online recommendation system for e-commerce based on apache mahout framework. Pages 153–158

[9] Shahab Saquib S., Jamshed S,and Rachid, A.Recommender Systems for E-commerce: In perspective of Business Strategies.

[10] Linda Hansson. « Product Recommendations in E-commerce Systems using Content-based Clustering and Collaborative filtering », 2015, atp10lh1@student.lu.se.

[11] Rahul S. Gaikwa, SandeeS. Udmale et Vijay K. Sambhe. « E-commerce Recommendation System Using Improved Probabilistic Mode »,2018 , https://www.researchgate.net/publication/320932136.

[12] Hijun Zhang, Gongwen Xu, et Pengfei Zhang. « Research on E-Commerce Platform-Based Personalized Recommendation Algorithm », 2016, School of Computer Science and Technology, Shandong Jianzhu University, Jinan, Shandong 250101, China.

[13] Samira Khodabandehlou. « Designing an e-commerce recommender system based on collaborative filtering using a data mining approach », 2019, Payame Noor University.

[14] Munercio Guia, Rodrigo Rocha Silva, et Jorge Bernardino. «A Hybrid Ontology-Based Recommandation System in e-Commerce », 2019.

[15] Samira Khodabandehlou.2019.Designing an e-commerce recommender system based on collaborative filtering using a data mining approach.Int. J. Business Information Systems, Vol. 31, No. 4.

[16] Liaoliang Jiang , Yuting Cheng , Li Yang , Jing Li , Hongyang Yan ,Xiaoqin Wang , 29 June 2018,A trust based collaborative filtering algorithm for E commerce recommendation system.

[17] Márcio Guia , Rodrigo Rocha Silva and Jorge Bernardino.2019. A Hybrid Ontology-Based Recommendation System in e-Commerce.

[18] Alfredo Cutolo , Giuseppe d'Aniello , Francesco Orciuoli , Francesca Pettinati ,Giuseppe Sansonetti Catello Vitagliano. An Ontology-Based Recommender System in E-Commerce.

[19] Analytics Vidhya.K-neighbors.[2021]. https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbors-algorithm-clustering/

[20] Javatpoint.[2021].https://www.javatpoint.com/k-nearest-neighbourhood-algorithm-for-mach

[21] Prateek Bajaj.20 Jun, 2018. Python-modules,Python. Creating linear kernel SVM in Python.[2021].https://www.geeksforgeeks.org/creating-linear-kernel-svm-in-python.

[22] Tejumade Afonja.Jan 2,2017.towards data science. Kernel Functions.[2021]. https://towardsdatascience.com/kernel-function-6f1d2be6091

[23] Programiz.Python User-defined Functions.[09-2021].
https://www.programiz.com/python-programming/user-defined-function

[24] IBM Tivoli Netcool/Impact, Version6.1.1.3. User-defined functions. [2021]. https://www.ibm.com/docs/SSSHYH_6.1.1/com.ibm.netcoolimpact.doc_6.1.1\\/common/dita/user_defined_functions_c.html.

[25] Teradata. What-is-Python. [2021].https://www.teradata.fr/Glossary/What-is-Python.

[26] Geeksforgeeks.Django.[2021].https://www.javatpoint.com/django-mvthttps://www.geeksforgeeks.org/django-project-mvt-structure/

[27] AnacondaDocumentations.[2021].
https://docs.anaconda.com/anaconda/navigator/index.html

[28] Tutorialspoint.nodjs.[2021].https://www.tutorialspoint.com/nodejs/nodejs_introduction.htm

[29] Business insider. csv file. [2021].https://www.businessinsider.fr/us/what-is-csv-file

[30] ClydeBank media. what is DB BROWSER. [2021]. https://www.clydebankmedia.com/definitions/technology/db-browser